5G enabled CAVs for Smart and Sustainable Mobility in Smart Cities

Soufiene Djahel, Yassine Hadjadj-Aoul

5G enabled CAVs for Smart and Sustainable Mobility in Smart Cities

Dr. Soufiene Djahel[1] & Dr. Yassine Hadjadj-Aoul[2]

[1]Manchester Metropolitan University, UK

[2]University of Rennes, INRIA, CNRS, IRISA, France

s.djahel@mmu.ac.uk, yassine.hadjadj-aoul@irisa.fr

IEEE ISC2 2020, September 28th , 2020

2020 IEEE International Smart Cities Conference

Outline

- ➢ The need for Smart Mobility
- ➢ How CAVs could support Smart Mobility
  - What is CAVs?
  - Its main characteristics
  - Main challenges
  - Potential solutions
- ➢ How 5G could support CAVs
  - Main caching challenges in 5G
  - Potential solutions
- ➢ Concluding remarks

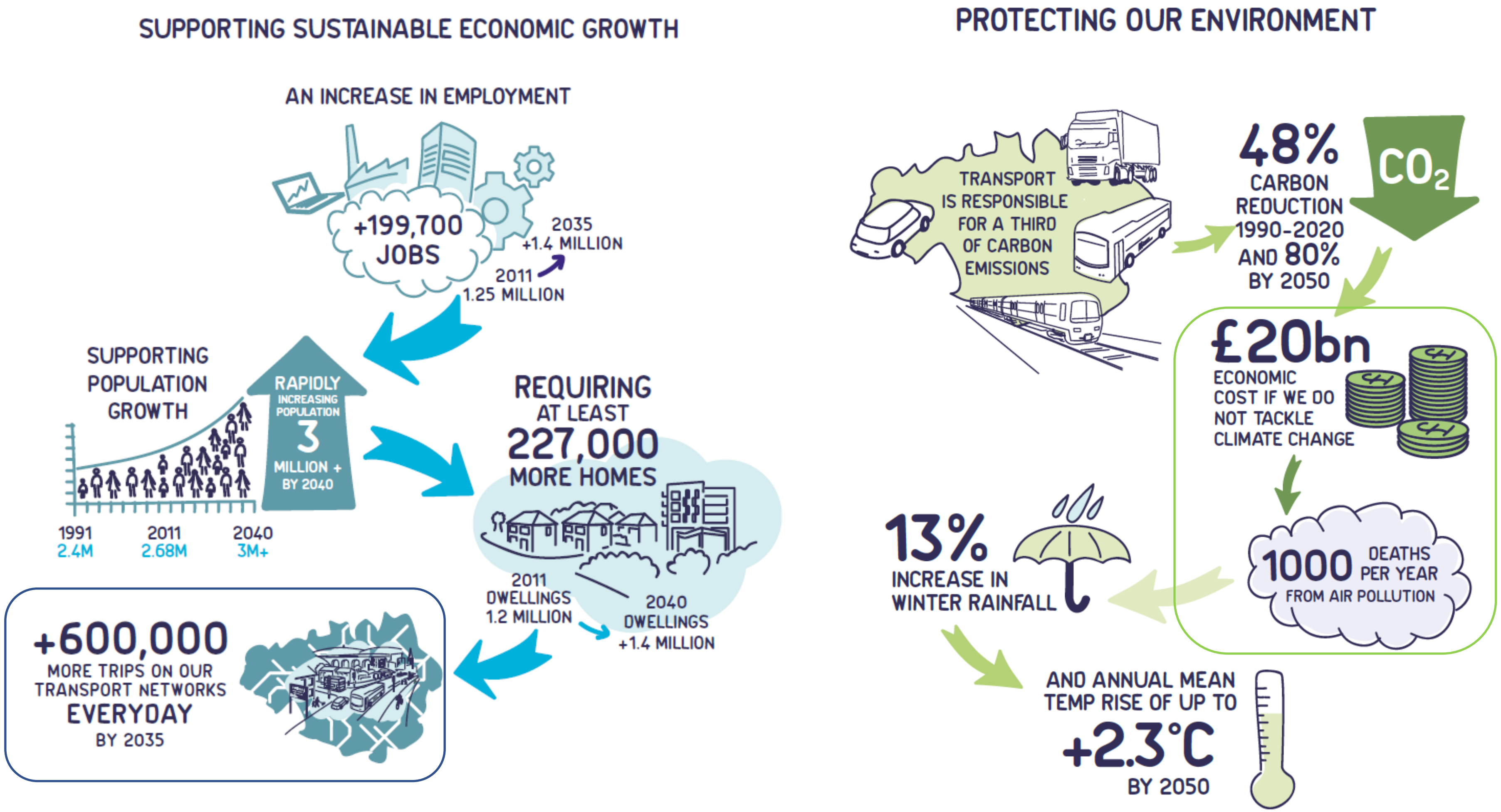
Transportation System

- ➢ Evolvable Critical System (ECS)
- ➢ Evolution:
  - Technology evolves, what to monitor evolves
  - Traffic pattern, volumes, complexity and data sources evolve

- ➢ Critical impact:
  - Congestion costs billions to world economy
  - 2013-2030: the economic loss due to traffic congestion in the UK alone is expected to be a staggering £307 billion *
  - => an increased congestion per-household cost from £1,426 in 2013 to £2,057 in 2030 *

* Source: INRIX

Road Transport in 2040: TfGM vision



Source: Simon Warburton talk at 'Moving North': Reducing transport emissions in the Northern Powerhouse workshop, April 12, 2018

Road Transport in 2040: TfGM vision

- New <u>lifestyles</u> and <u>Travel</u> Attitudes

- 77% of 18-35 year olds plan to live in urban centres – “in vibrant, compact and walkable communities full of economic, social and recreational activities.” (2010, Brookings Institution)

- Hence, the need for New Mobility Models

Source: Simon Warburton talk at 'Moving North': Reducing transport emissions in the Northern Powerhouse workshop, April 12, 2018

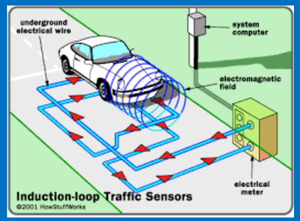
New Mobility Models in Smart Cities Era

- Make the mobility of people and goods worldwide:
  - Faster
  - Cheaper
  - Eco-friendly
  - and safer

- Mobility as a Service (MaaS)

Efficient Traffic Data Collection & Analytics

Induction loops, Sensors, IoT Sensors, Crowd Sensing, CCTV cameras, etc.



Traffic Congestion Reduction

ICT driven solutions, Prediction & mitigation techniques, Green route for emergency vehicles, etc.

Connected & Autonomous Vehicles

Safety and Traffic Efficiency, & Infotainment Applications

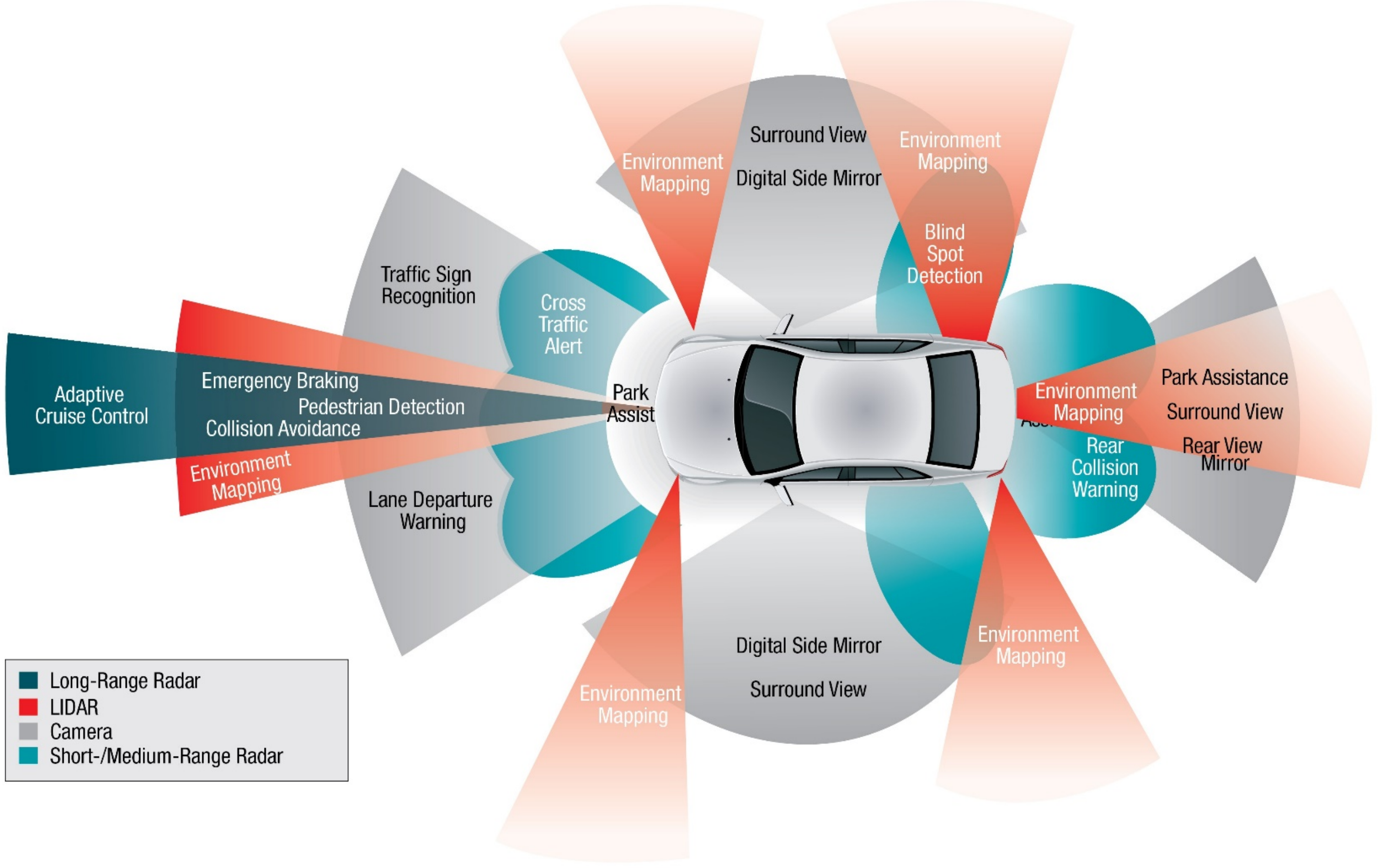
Mobility as a Service (MaaS)

Autonomous Vehicles

Autonomous (Self-driving) Vehicles (1/3)

- Localization: where am I?
- Sensing the surroundings: what’s happening around me?
- Perception (fusion of sensor data)
- Reasoning and decision making
- Motion control

Autonomous (Self-driving) Vehicles (2/3)



Autonomous (Self-driving) Vehicles (3/3)

GPS

DSRC
(760 MHz, 5.9 GHz)

Cellular technology
(LTE, 5G, etc.)

Bluetooth

Electronic Toll
Collection

Wi-Fi

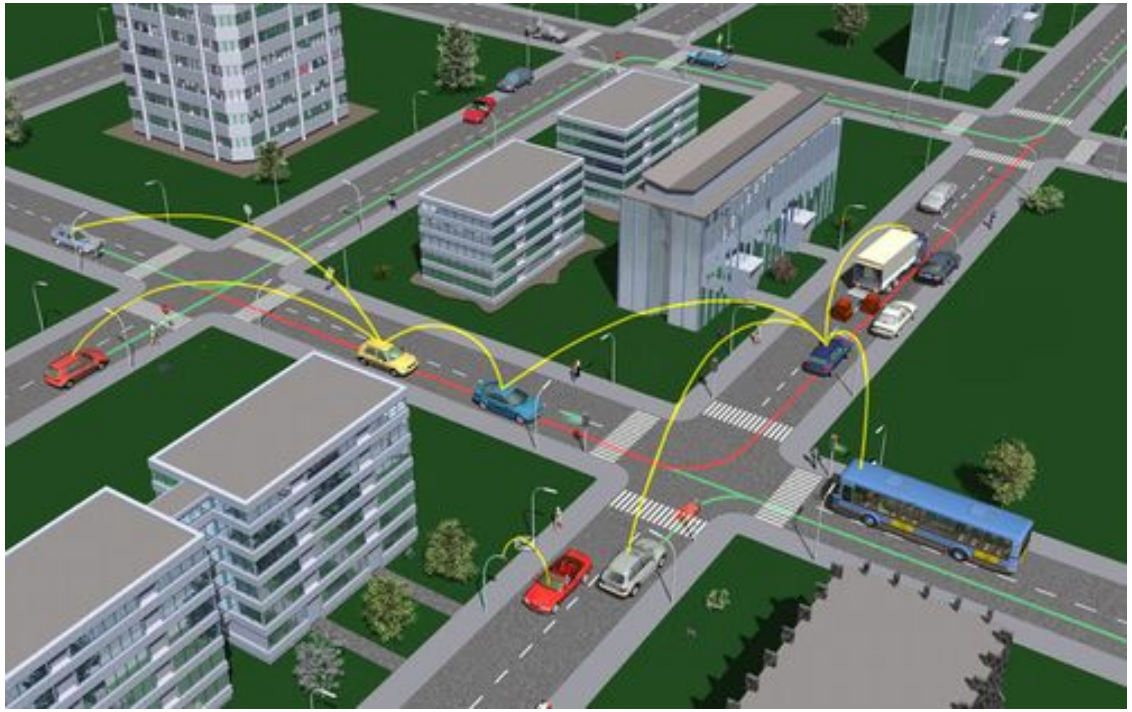
Radar

Wireless Battery
Charging

Visible Light Communications
(VLC)

Connected & Autonomous Vehicles (CAVs)

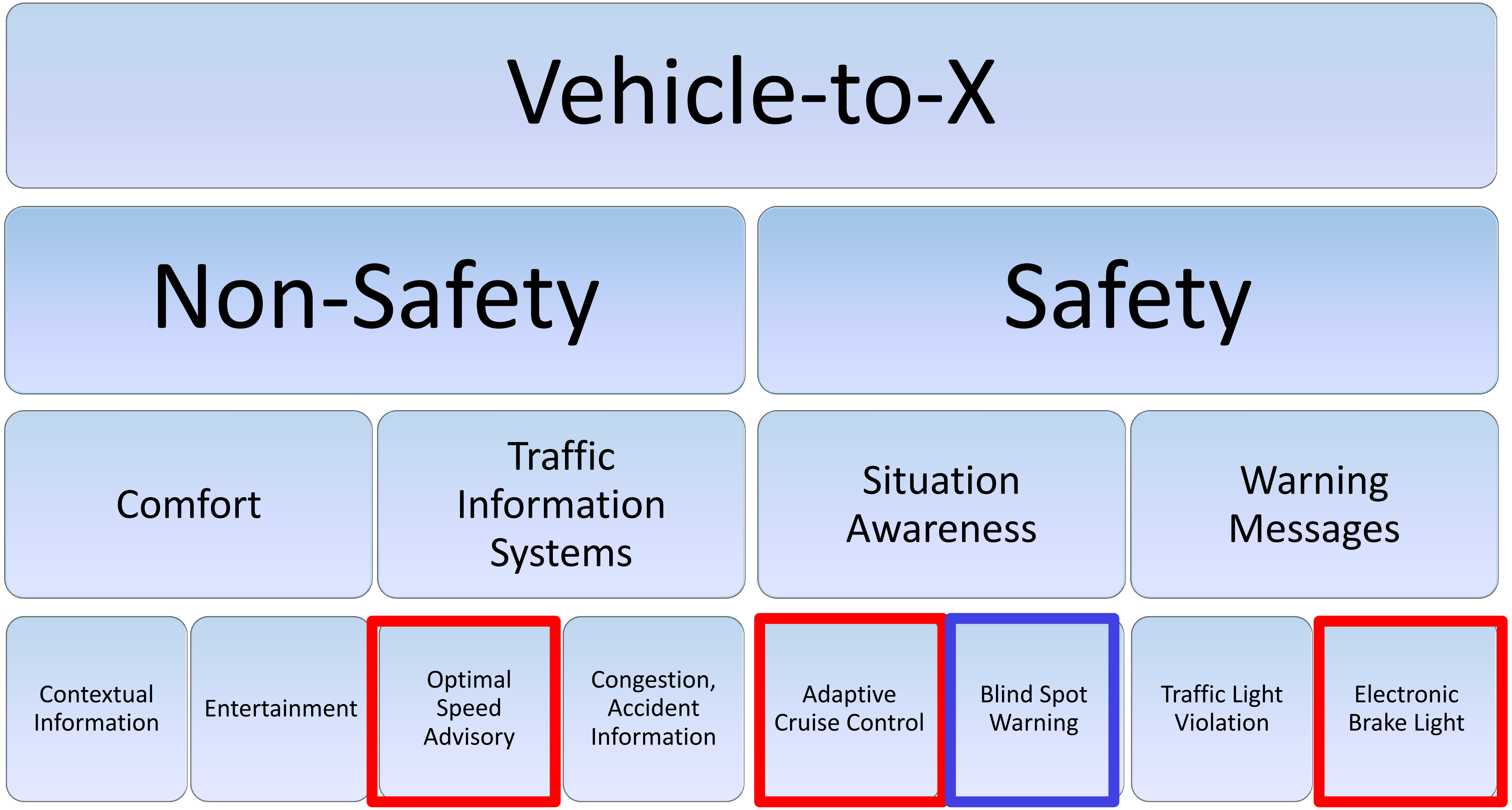
Vehicular Networks



- Wireless communication between
  - Vehicles (V2V)
  - Vehicles and the road infrastructure (V2I)
  - Vehicles and X?

- Promise to solve many of today's road traffic problems by
  - Improving road users' safety
  - Shortening their trip times
  - Enhancing their travel experience
  - Reducing air pollution

Taxonomy of use cases



C. Sommer and F. Dressler, Vehicular Networking, Cambridge University Press, 2014

Taxonomy of use cases

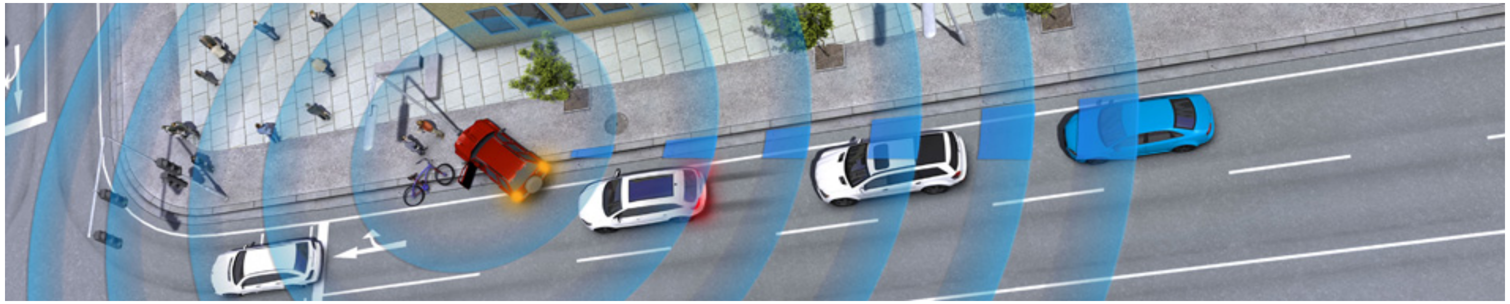
Vehicle-to-X

Non-Safety	Safety
Many messages	Few messages
High data rate	Small packet size
Low latency demands	High latency demands
Low reliability demands	High reliability demands

C. Sommer and F. Dressler, Vehicular Networking, Cambridge University Press, 2014

Potential benefits

1 - Safer driving



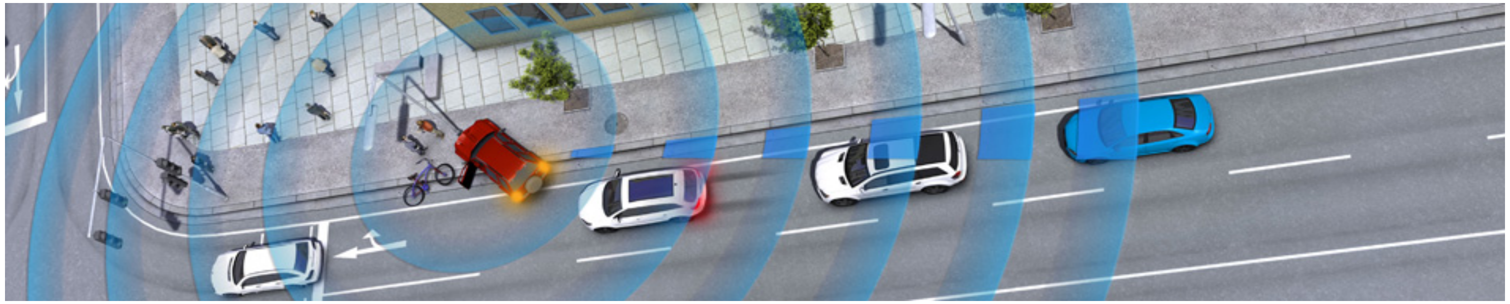
- Safety information sharing and dissemination[1]
- Cooperative manoeuvres
  - Safe and efficient lane change[2]
- Hidden obstacles notification
- …

[1]S. Zemouri, **S. Djahel**, and J. Murphy. ***A Fast, Reliable and Lightweight Distributed Dissemination Protocol for Safety Messages in Urban Vehicular Networks***. Ad Hoc Networks Elsevier, Vol. 27, pp. 26-43, 2015.

[2]A. Lissac, **S.Djahel** and J. Hodgkiss. ***Infrastructure Assisted Automation of Lane Change Manoeuvre for Connected and Autonomous Vehicles***. IEEE ISC2, Oct. 14-17, 2019.

Safer driving - Related challenges

- ➢ How to ensure fast & reliable safety information dissemination?
- ➢ How to avoid periodic beacons congestion?
  - At microscopic level
  - At macroscopic level



Beacons congestion control

- Safety and critical information (periodic broadcast)
  - Inter-vehicular communication (infrastructure-free)
  - Urban scenarios
- Two types of messages
  - Event driven warning messages
  - Control messages (beacons, , CAMs, etc.)

➔ **Wireless congestion problem**

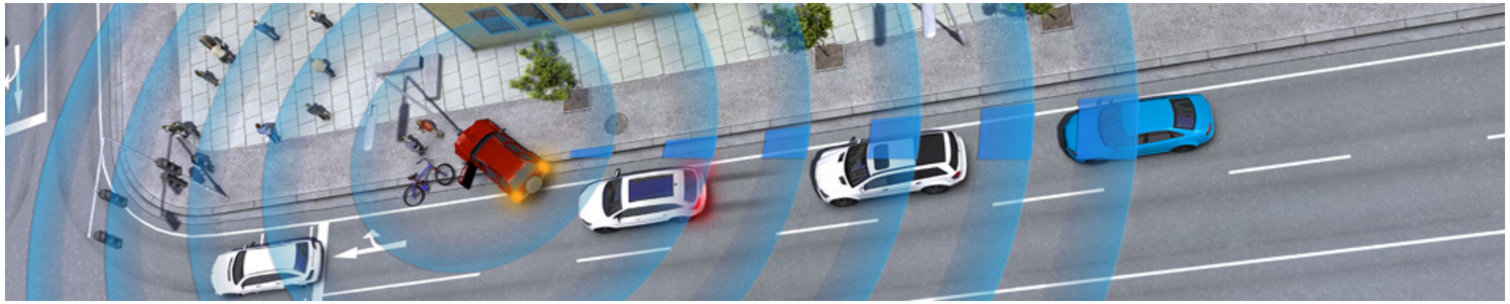
Ph.D. thesis "**Enhanced Wireless Congestion Control in Urban Vehicular Environments**", Sofiane Zemouri, University College Dublin, Ireland , July 2016

Safety messages dissemination

- Critical information, multi hop transmission, high periodicity, limited timeframe, short to medium length messages

- Goal, limit the amount of data inserted in the network
  - Get rid of unnecessary retransmissions after failures
  - Channel condition (shadowing, Doppler Effect,…), packet collisions, integrity of the message, sender receiver distance

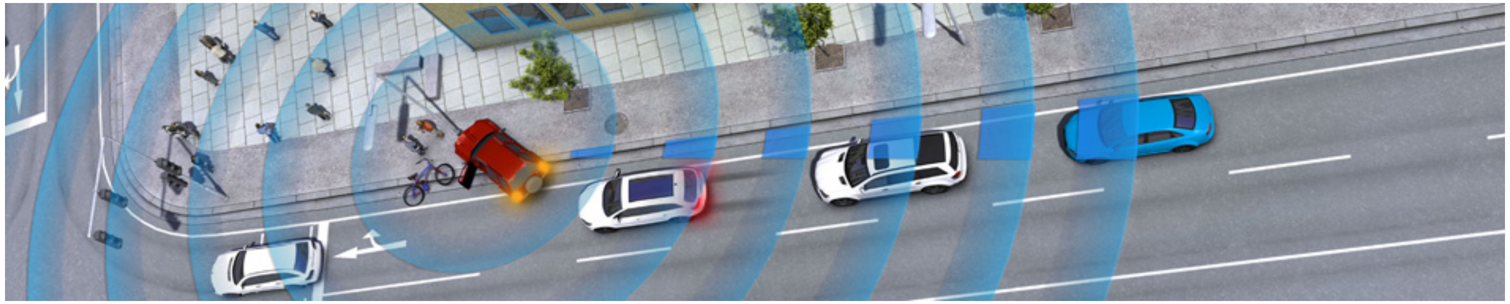
➔ Who’s the most suitable next forwarder?



Safety messages dissemination

Road-Casting Protocol (RCP)

- Distance factor (D)
  - A function of the distance between the sender & the potential next forwarder, the next junction and the potential next forwarder, and the next junction to the transmission range of the sender
- Link Quality factor (LQ)
  - A function of the signal quality (SQ), the channel quality (CQ) and the collision probability (CP)



S. Zemouri, **S. Djahel**, and J. Murphy. *A Fast, Reliable and Lightweight Distributed Dissemination Protocol for Safety Messages in Urban Vehicular Networks*. Ad Hoc Networks Elsevier, Vol. 27, pp. 26-43, 2015

Safety messages dissemination

Road-Casting Protocol (RCP)

- Retransmission Probability

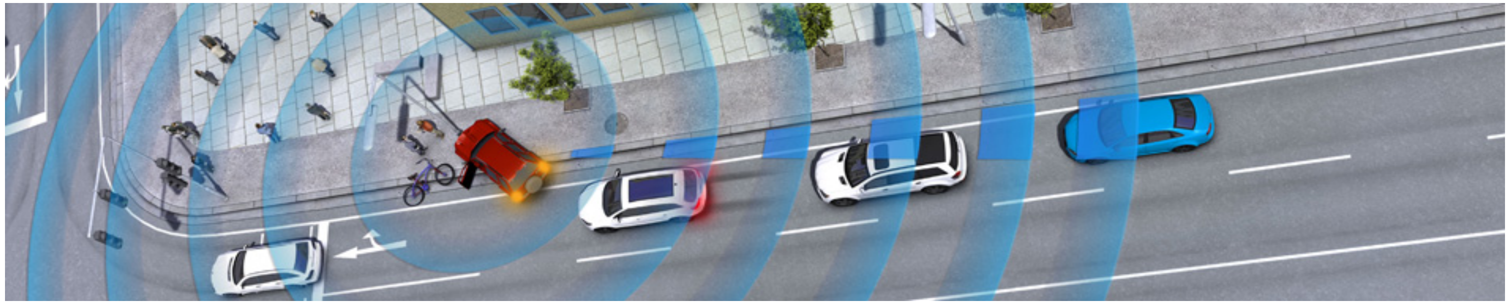
$$P = (1 - \omega_P)D + \omega_P LQ$$

$0.5 \leq \omega_P \leq 1$
D: Distance Factor
LQ: Link Quality Factor

- Retransmission Waiting Time (Backoff)

$$WT = CW \times (1 - P) + \delta$$

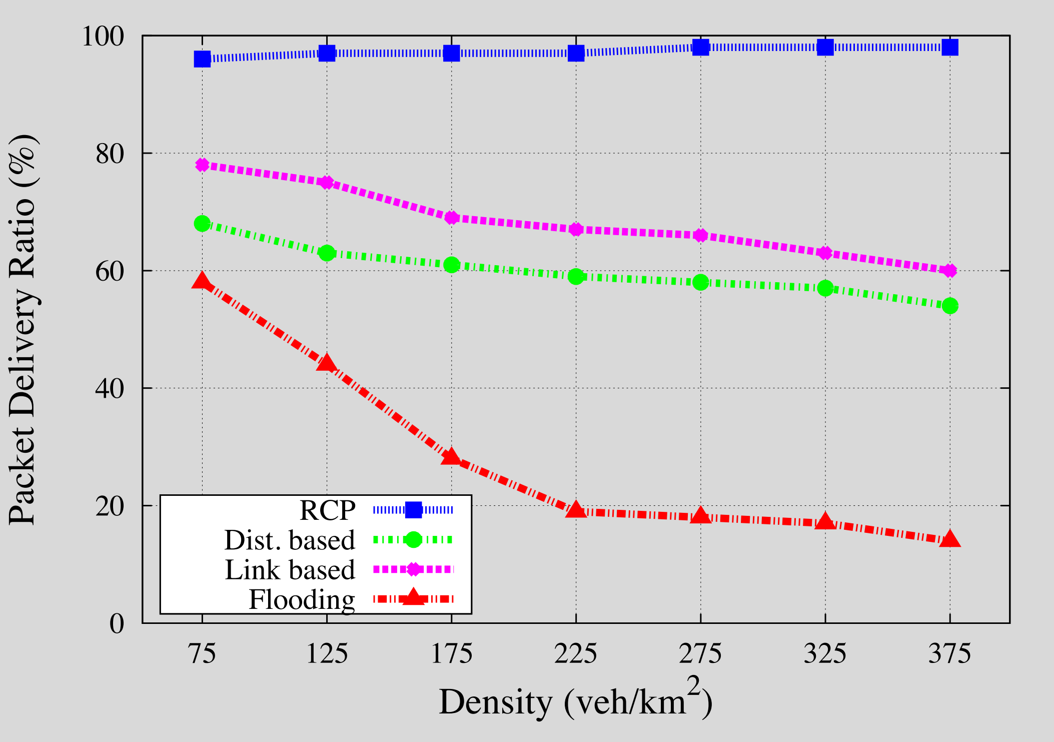
δ: Random value in µs so that $2\mu s < \delta < 9\mu s$
CW: Contention Window

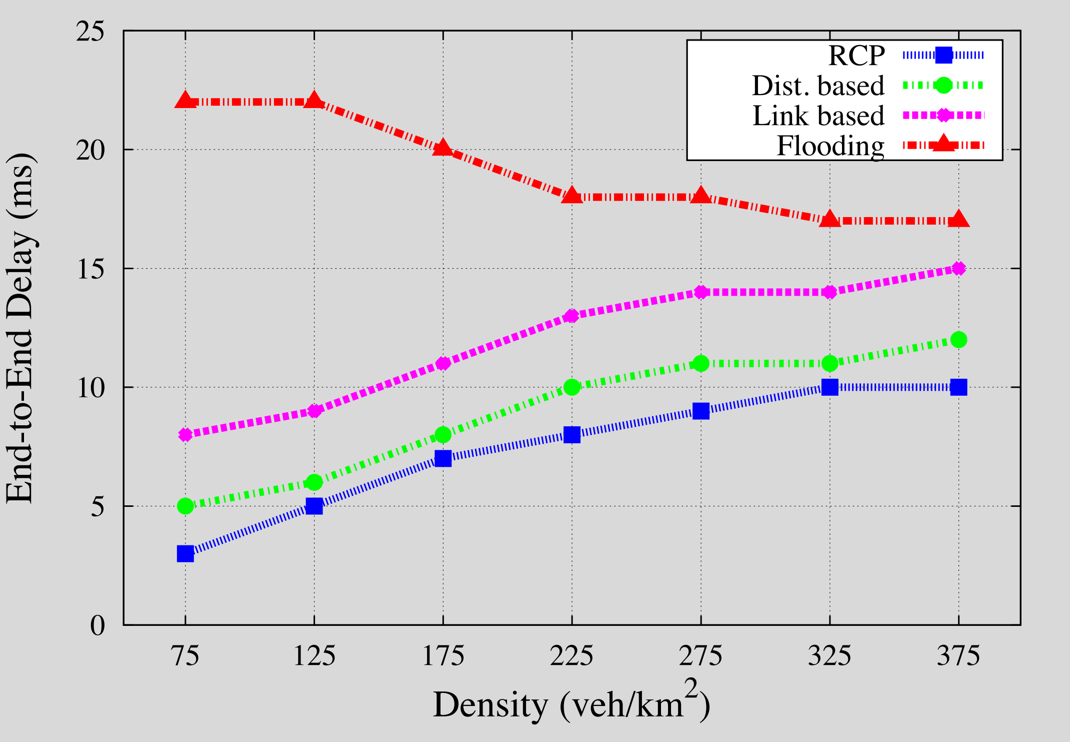


Evaluation results

Bit rate: 6Mb/s
CW: [15,1023]

Data message size: 100 bytes
Broadcast frequency: 1hz



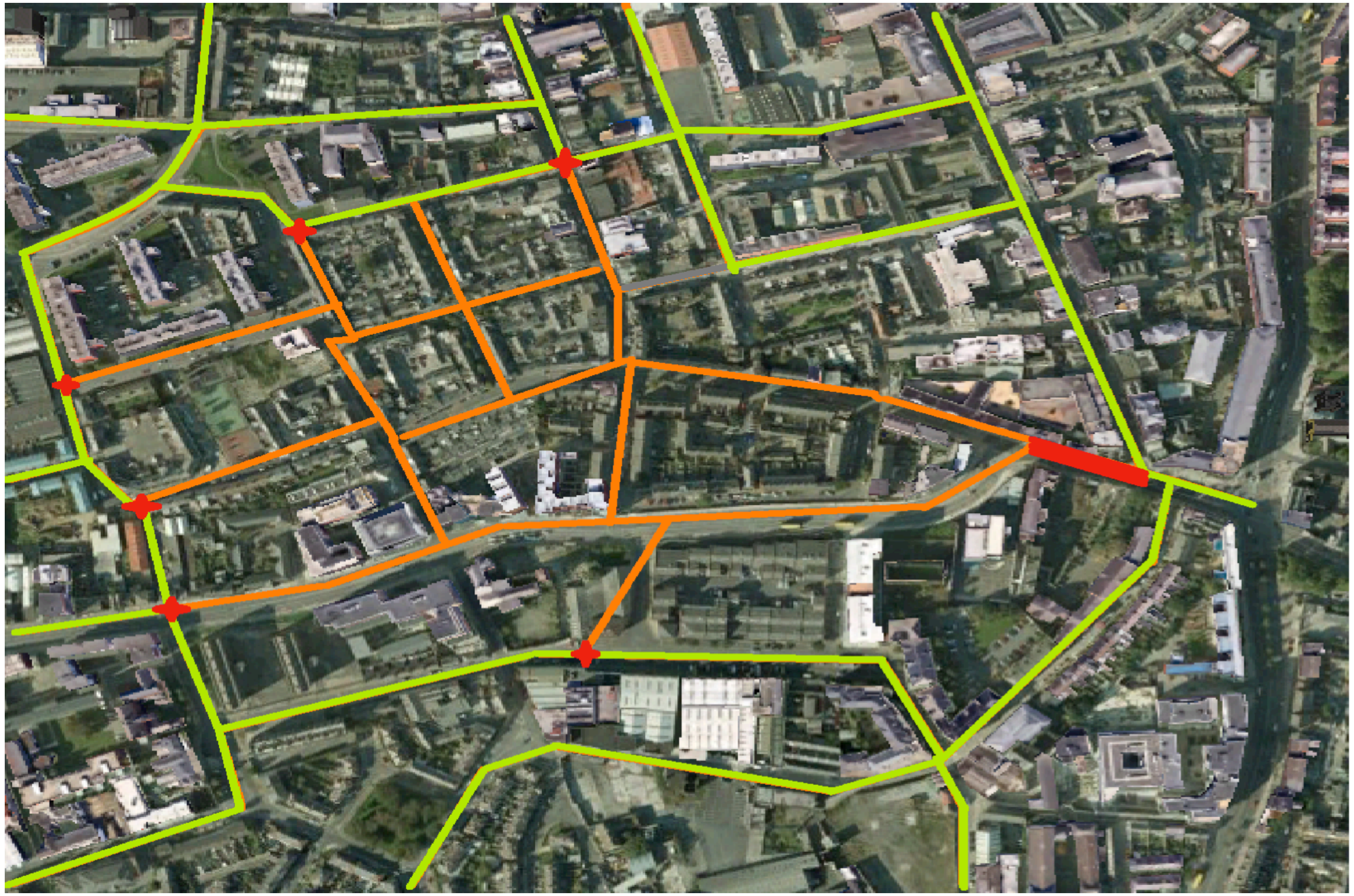


Impact of vehicles density on Packet Delivery Ratio and End-to-End Delay: RCP vs. distance based forwarding, link based forwarding and flooding

Safety messages dissemination

- Microscopic congestion control ✓
  - Amount of data inserted in the network on a point-to-point level
- Macroscopic congestion control
  - The extent of data inserted in the network (end-to-end)
- Determine the Region of Interest (RoI) for multi-hop safety messages
- Keep the amount of data inserted in the network minimal

Safety messages dissemination

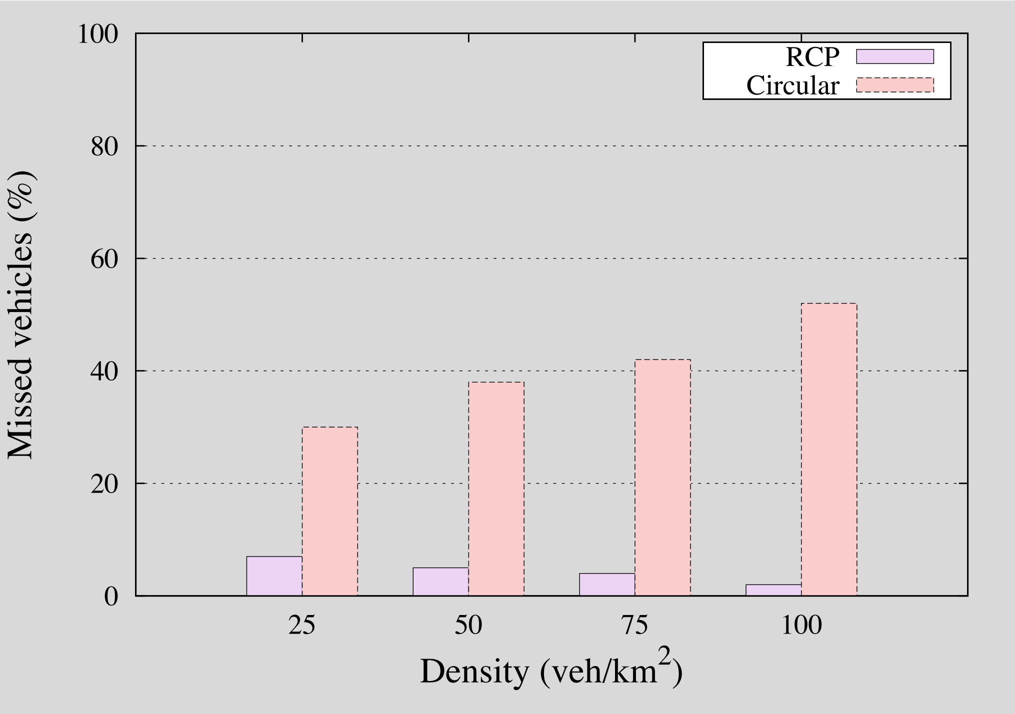


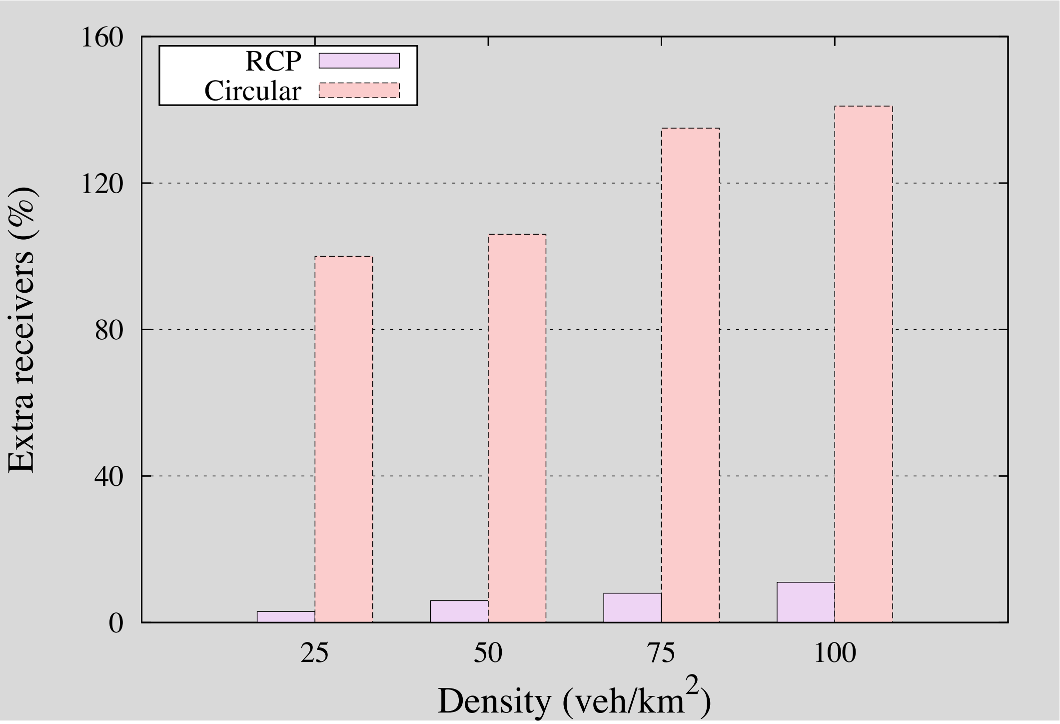
The set of critical junctions (red ticks) delimitating the zone where vehicles are already blocked due to an incident

Safety messages dissemination

Road map connectivity check performed by each receiver of an emergency message: using simplified A* algorithm

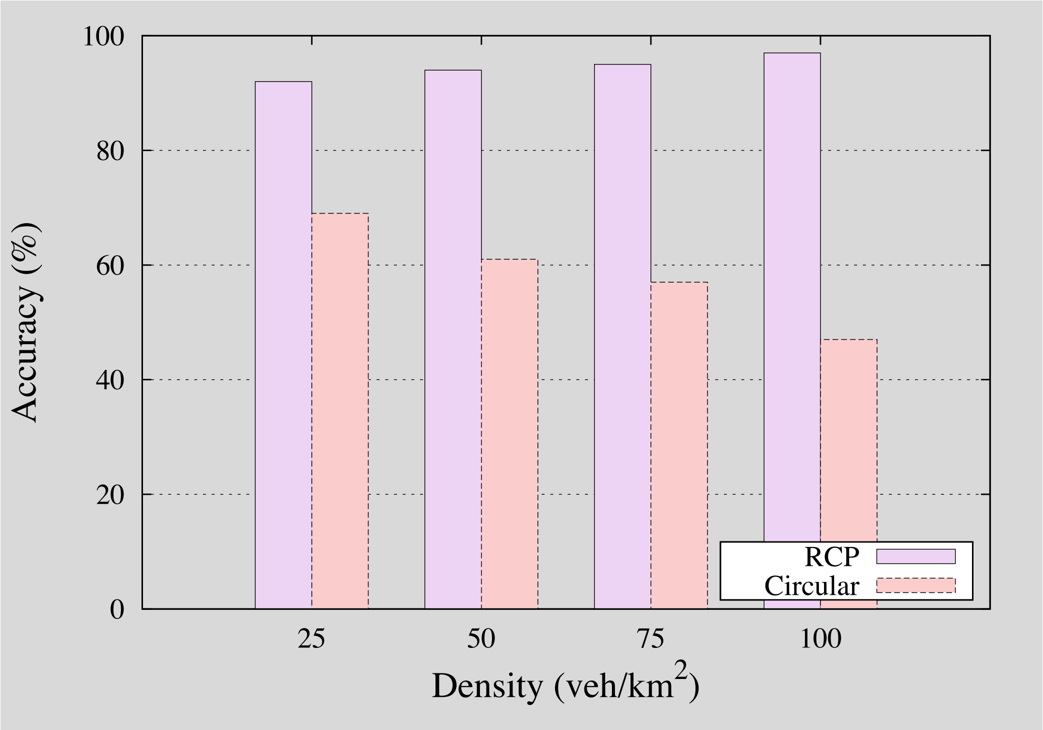
Evaluation results

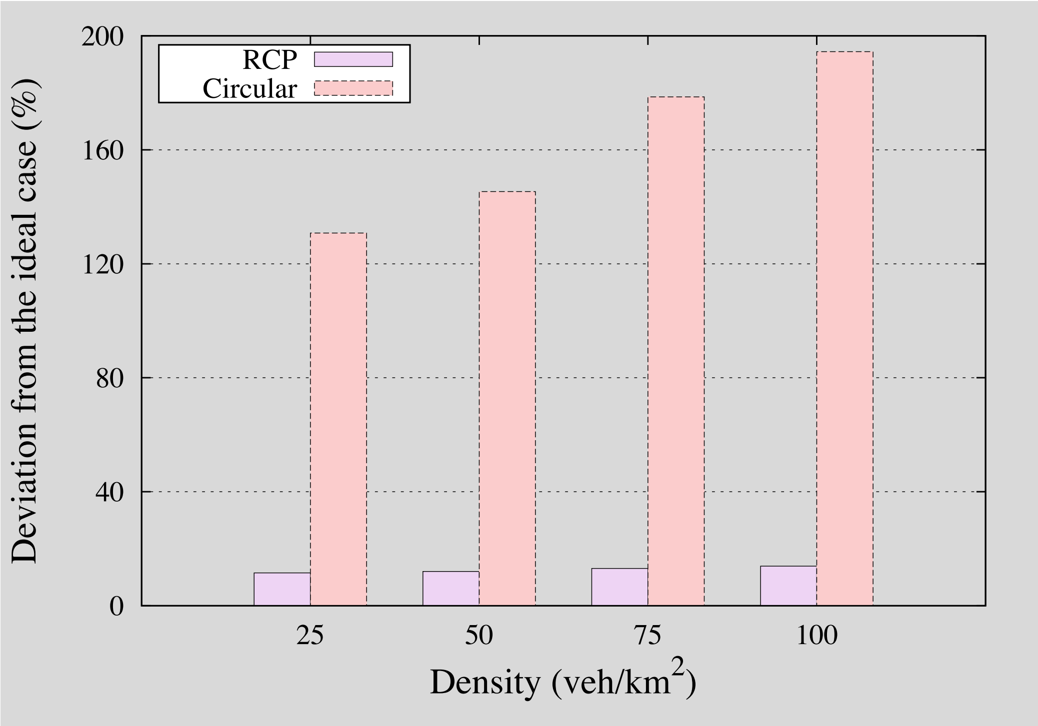




The ratio, under different vehicle densities, of vehicles located inside and outside the RoI which did not receive the emergency message (Algiers map)

Evaluation results





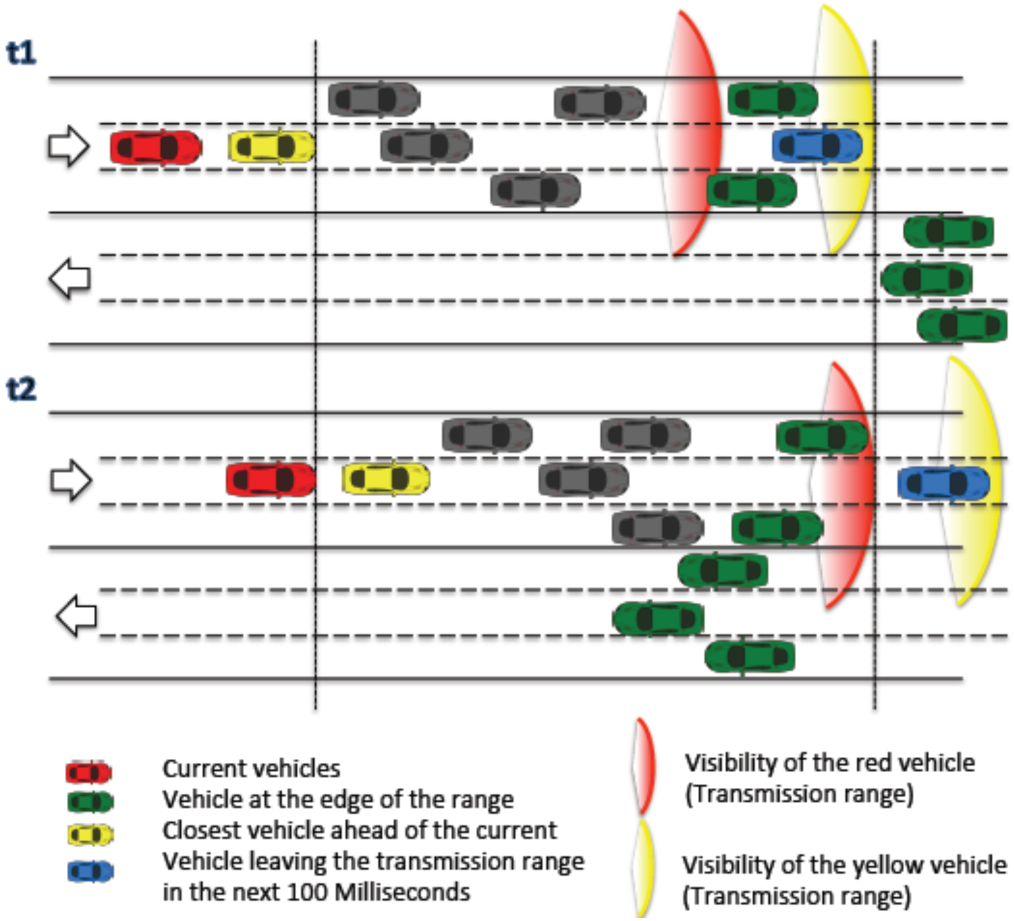
Impact of vehicles density on the accuracy of the RoI definition: RCP scheme RoI vs. circular RoI
(Algiers map)

Congestion control for periodic beacons

Altruistic prediction mechanism

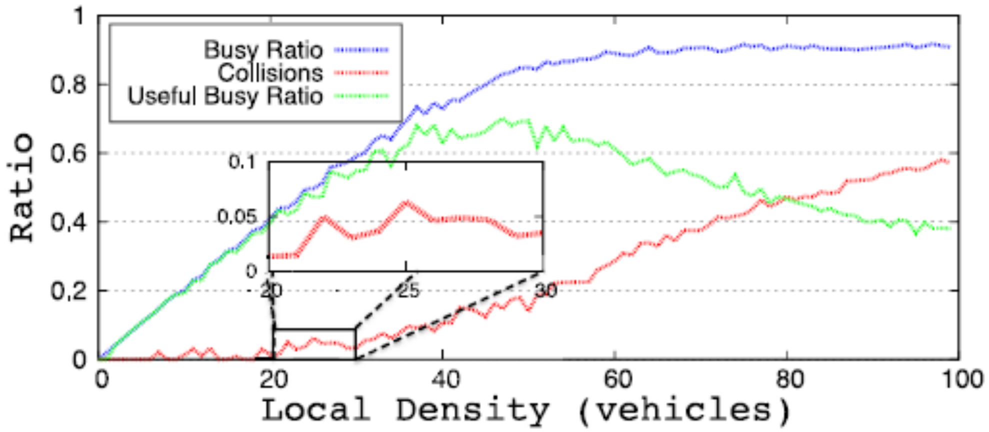
Altruistic prediction mechanism

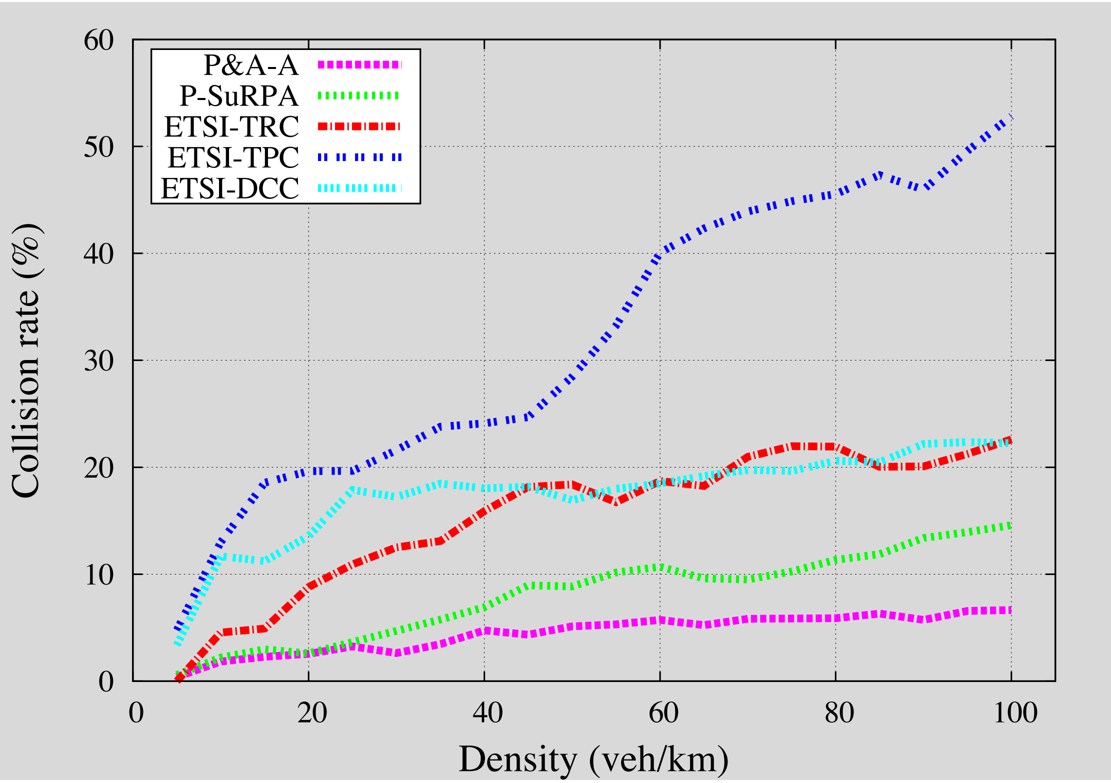
Prediction & Adaptation – Algorithms (P&A-A)

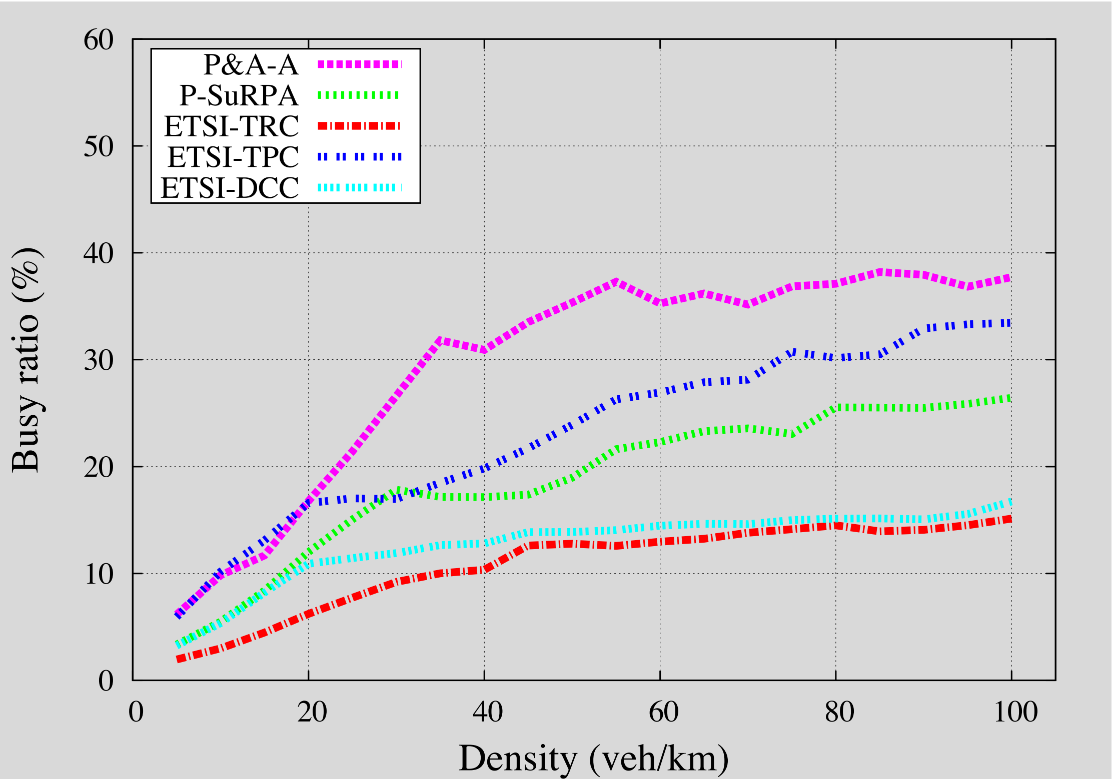


S. Zemouri, **S. Djahel**, J. Murphy. An Altruistic Prediction-Based Congestion Control for Strict Beaconing Requirements in Urban VANET. in IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2019

P&A-A evaluation results







Impact of vehicles density on the achieved channel busy ratio and collision rate:
P&A- A vs. **SuRPA** and **ETSI** schemes (Kirchberg map)

Summary

- Road-Casting Protocol (RCP)
  - Next forwarder selection (microscopic congestion control)
  - Region of Interest definition (macroscopic congestion control)
- Prediction & Adaptation – Algorithms (P&A-A)
  - Altruistic prediction mechanism
  - Successive transmission Rate and Power adaptation

2- Faster & more comfortable journeys

- Traffic congestion and accidents information sharing
- Support for faster travel and emergency service delivery
  - Adaptive Traffic Lights Control (TLC) systems[1]
- More efficient re-routing[2]

[1]ALEKO, D.R.; **Djahel, S**. **An Efficient Adaptive Traffic Light Control System for Urban Road Traffic Congestion Reduction in Smart Cities**. *Information* 2020, *11*, 119.
[2]Ph.D. thesis , "**Reducing Non-Recurrent Urban Traffic Congestion using Vehicle Re-routing**", Shen Wang, Dublin City University, Ireland, July 2016

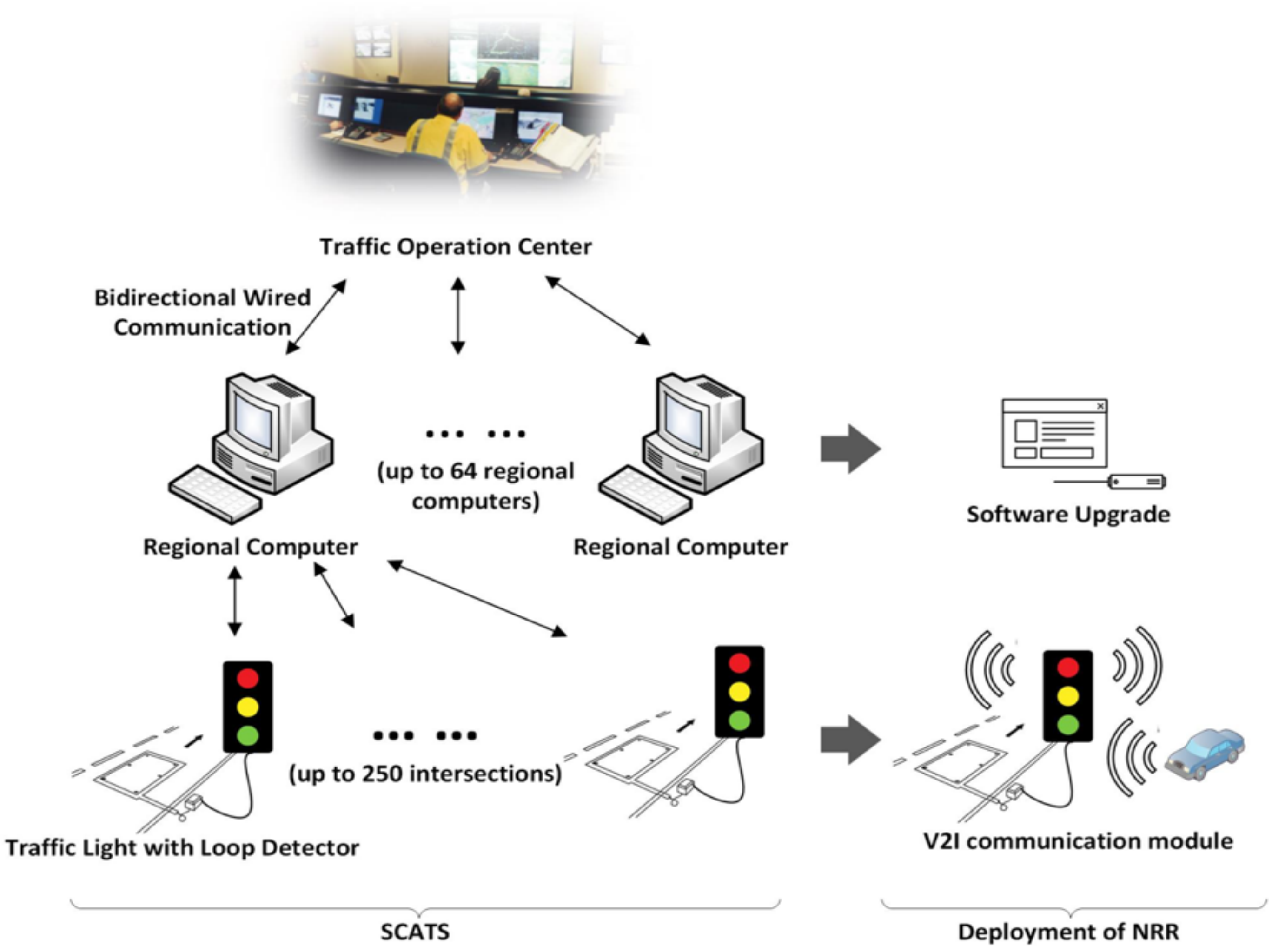
Faster journeys - Related challenges

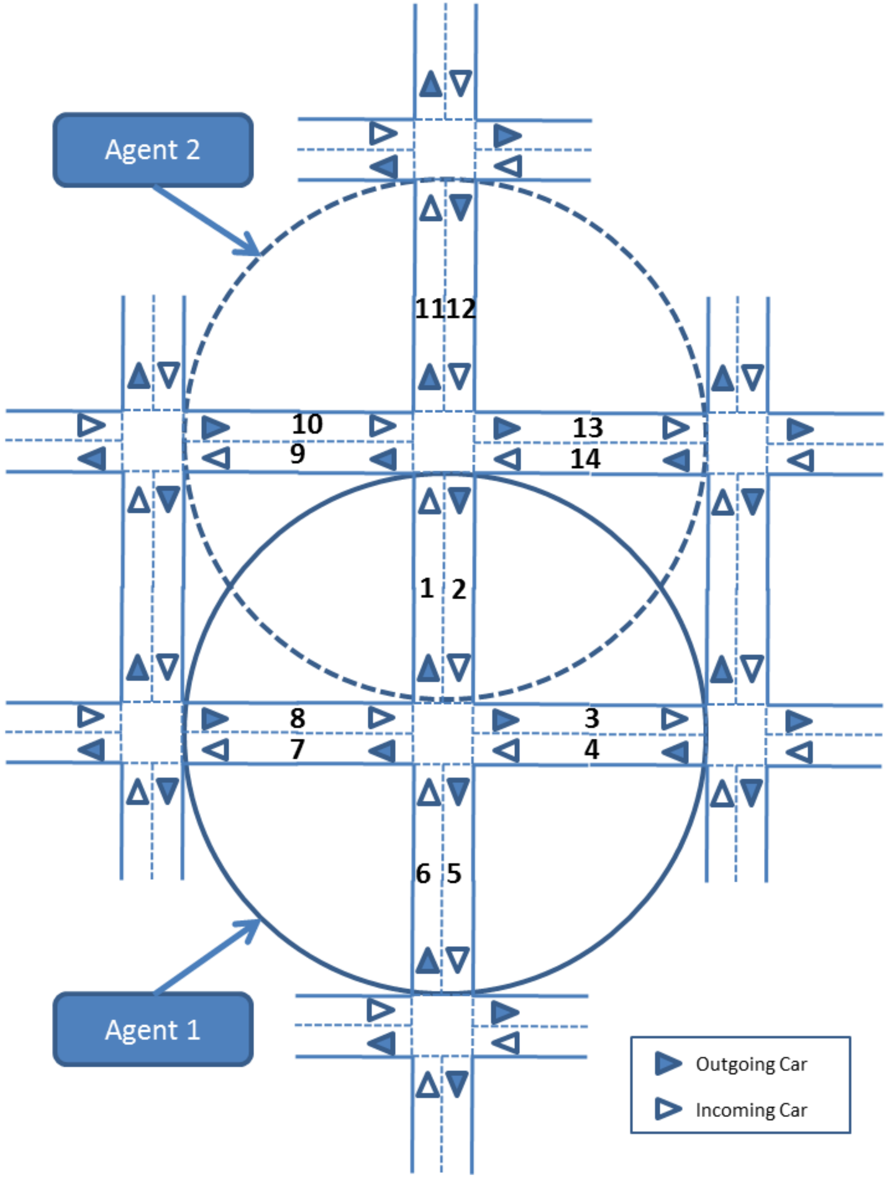
- How to design efficient re-routing strategies to efficiently deal with unexpected events on the road

- i.e., re-routing protocols that minimize the extra delay caused by the unexpected event

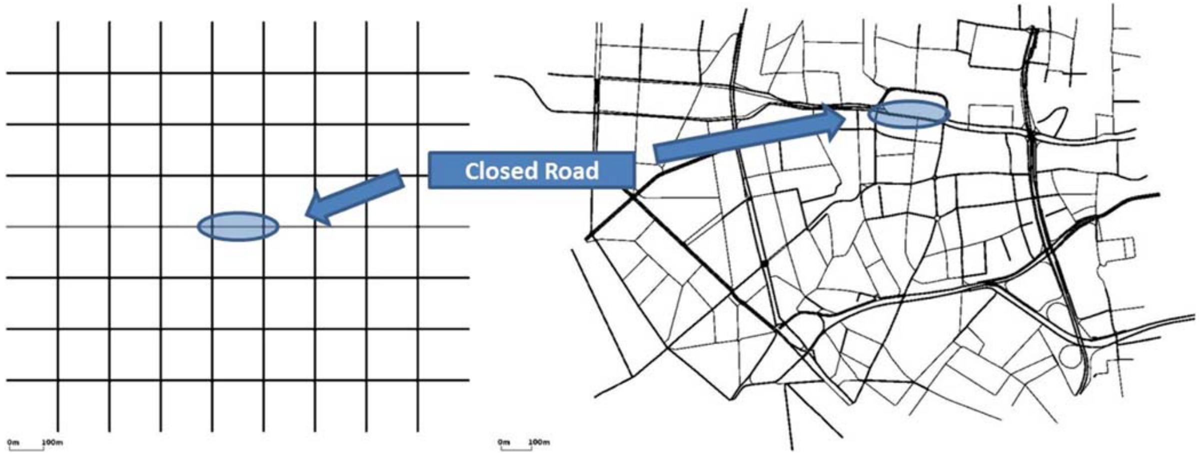
- Answer: Next Road Re-routing (NRR)

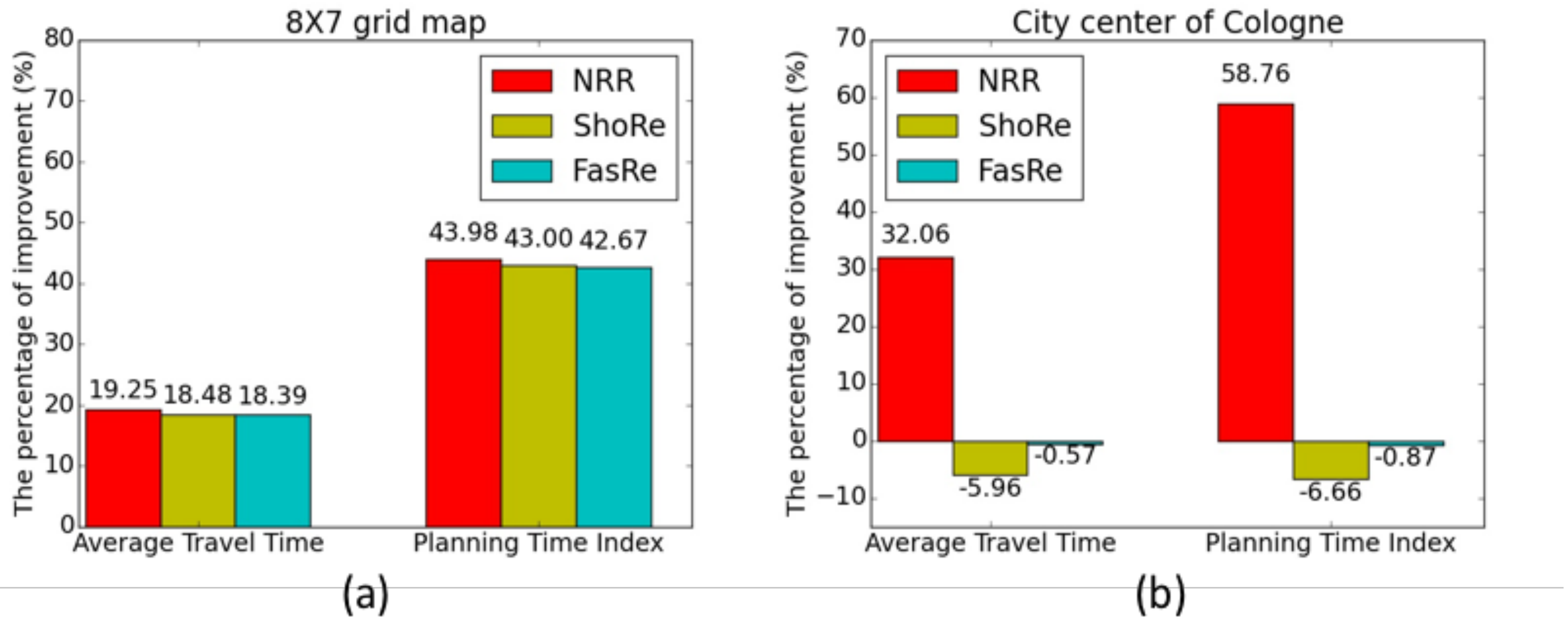
Next Road Re-routing (NRR)





NRR's evaluation results

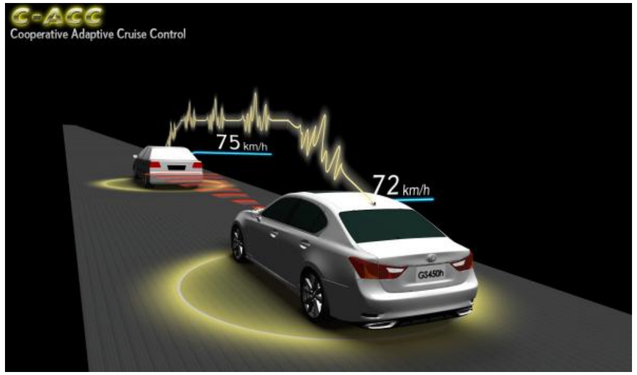




S. Wang, **S. Djahel**, Z. Zhang, and J. McManis. Next Road Rerouting: A Multi-Agent System for Mitigating Unexpected Urban Traffic Congestion. in IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 10, pp. 2888-2899, Oct. 2016

3 – Clean or Green mobility

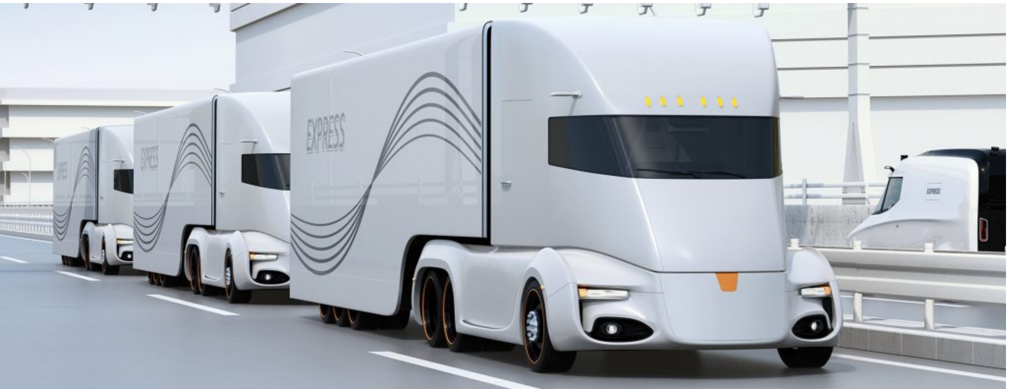
- Support the adoption of Electric Vehicles (EVs)
- <u>Cooperative</u> Adaptive Cruise Control (CACC)



- Cars share info on speed changes in real-time
  - More efficient Adaptive cruise control system
  - Reduced carbon emissions

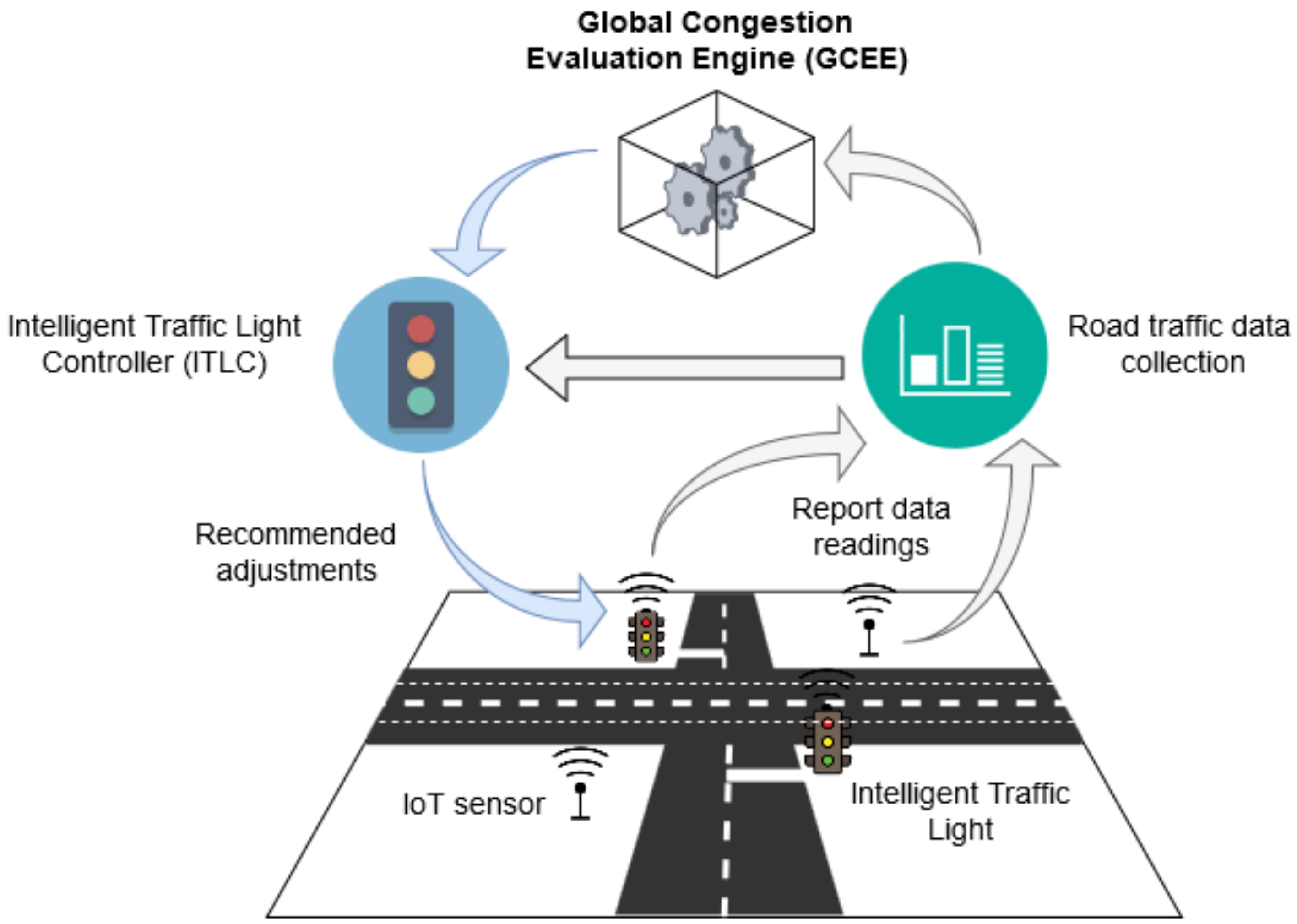
CACC – Related challenges

- Truck platooning requires new algorithms to determine:
  - Optimal frequency of platooning and length of a platoon
  - Smooth and safe integration of platoons in the surrounding traffic
    - Optimal speed
    - Gaps
    - Distance to surrounding traffic
- Evaluating its impact on the infrastructure

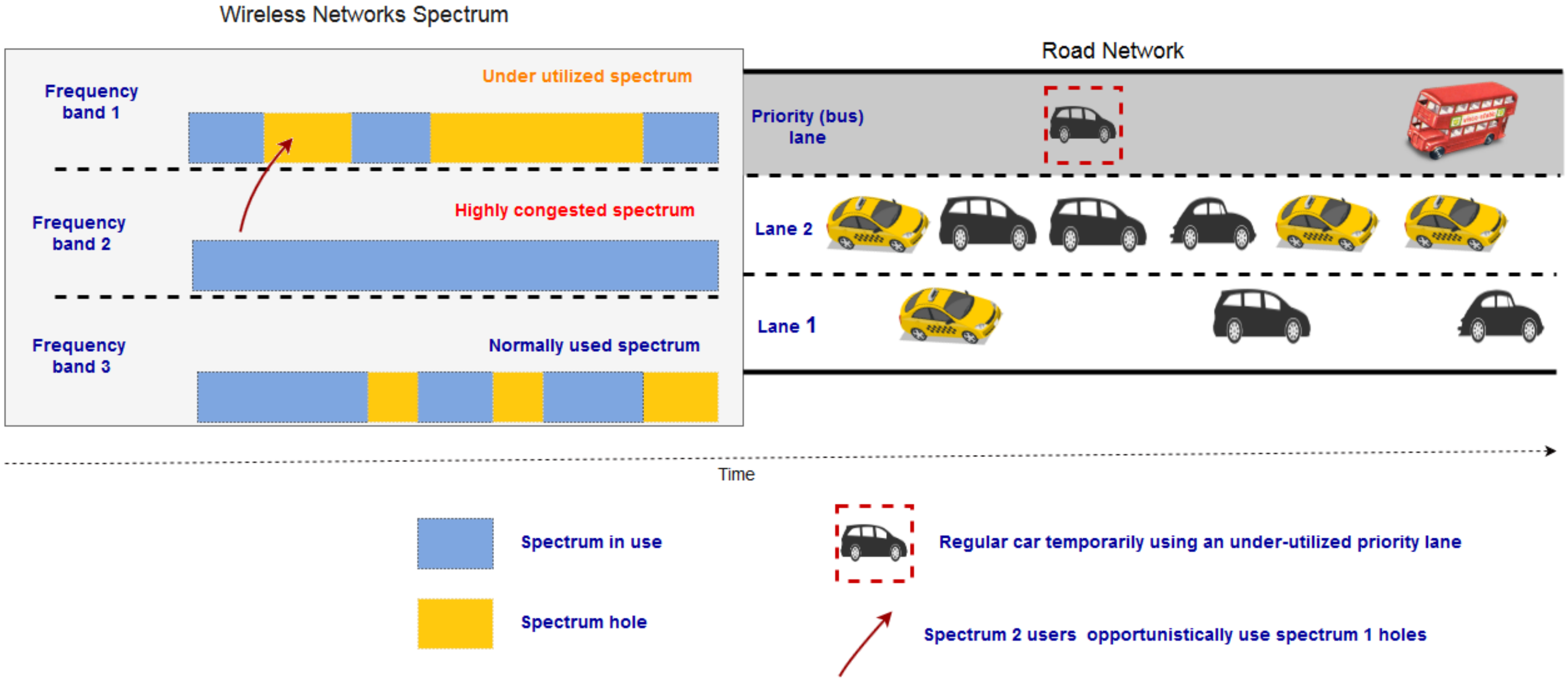


4 – Traffic congestion reduction

- CR-TMS



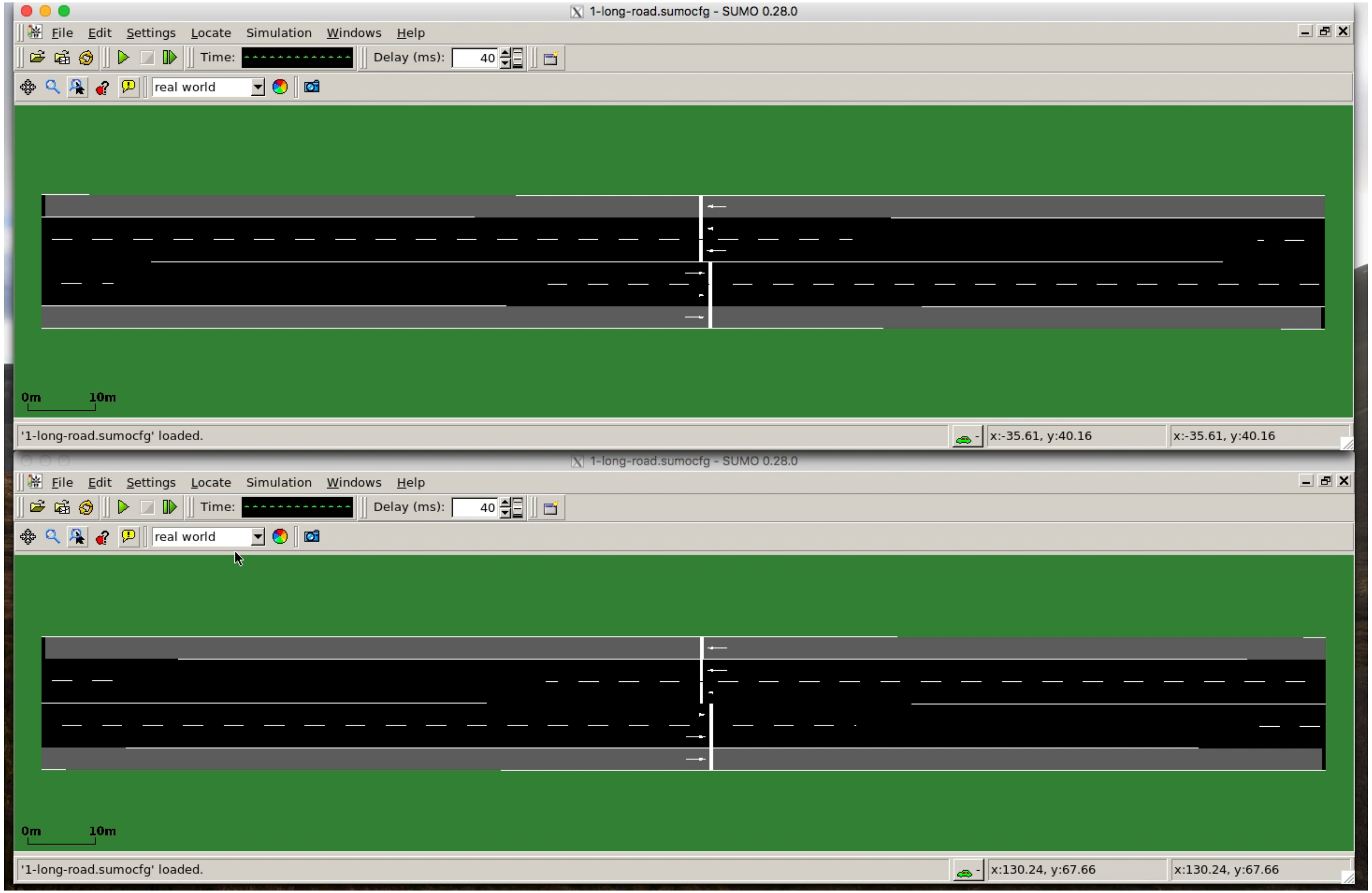
CR-TMS - key principle



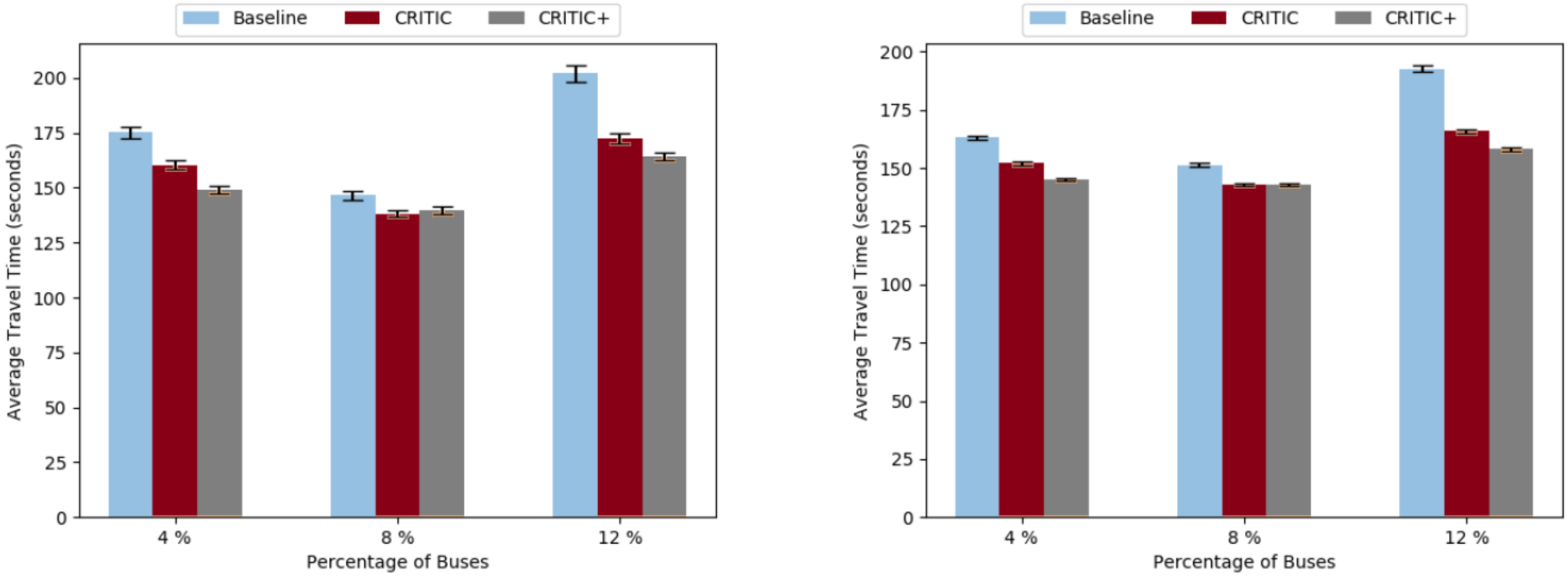
$$\Pr(i) = (W_{power} \times Power_{value}(i)) + (W_{trip} \times Trip_{value}(i)) + (W_{waiting\text{-}time} \times Waiting\text{-}time_{value}(i))$$

S. Djahel, Y. Hadjadj-Aoul and R. Pincemin. CR-TMS: Connected Vehicles enabled Road Traffic Congestion Mitigation System using Virtual Road Capacity Inflation. The 23rd IEEE International Conference on Intelligent Transportation Systems (Virtual Conference), Sep. 20 – 23, 2020.

CR-TMS demo


1-long-road.sumocfg - SUMO 0.28.0
File Edit Settings Locate Simulation Windows Help
Time:
Delay (ms): 40
real world
0m 10m
'1-long-road.sumocfg' loaded.
x:-35.61, y:40.16
x:-35.61, y:40.16
1-long-road.sumocfg - SUMO 0.28.0
File Edit Settings Locate Simulation Windows Help
Time:
Delay (ms): 40
real world
0m 10m
'1-long-road.sumocfg' loaded.
x:130.24, y:67.66
x:130.24, y:67.66

CR-TMS – evaluation results



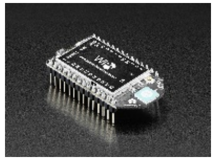
(c) Electric vehicles (d) Private vehicles

Impact of the percentage of buses on the achieved ATT
7*7 grid map with 25% of bus lanes

5 – Improved ITS services

Vehicle as a Resource
VaaR

Sensing



Computing

Data relay

Data storage

Infotainment

Localisation

VaaR – Related challenges

➢ How to ensure efficient resources provision to ITS applications?

- Rapid and dynamic change of resources availability due to vehicles mobility
- Modeling and accurately predicting users’ request patterns
- Handover between clusters of CAVs service providers

➢ Solution: AI enabled techniques such as Federated learning

Emerging Challenges

Emerging challenges for CAVs

- Bandwidth scarcity of the DSRC spectrum
  - TV White Space? LTE-A/5G/6G?
- Congestion and awareness control
  - Microscopic & Macroscopic – Machine learning
- Electric vehicles (EVs)
  - Route planning & re-routing, charging planning
- Security, privacy and liability
  - Cyber-attacks, location privacy, crash, etc.

Emerging challenges for CAVs on the Sky

- CAVs on the Sky = UAVs or Drones
  - Flexible and low cost solution with modest infrastructure needs
- Ground to Air and Air to Air communication
- Key challenges for cooperative UAVs
  - Optimal routing and recharging scheduling
  - Lightweight dynamic re-planning
  - Optimal tasks redistribution strategies
- CAVs assisted cooperative UAVs?
- UAVs assisted CAVs?

Conclusion

- CAVs is an emerging technology facing many challenges & offering many opportunities
- Successfully overcoming these challenges promises a significant societal and economic impact
- Most experts foresee that CAVs will be a key enabler of Smart Mobility, thus addressing its major technical challenges is essential

Outline

- The need for Smart Mobility
- How CAVs could support Smart Mobility
  - What is CAVs?
  - Its main characteristics
  - Main challenges
  - Potential solutions
- How 5G could support CAVs
  - Main caching challenges in 5G
  - Potential solutions
- Concluding remarks

Focus:
Efficient caching in 5G and Beyond 5G (B5G)

Dr. Soufiene Djahel[1] & Dr. Yassine Hadjadj-Aoul[2]

[1]Manchester Metropolitan University, UK

[2]University of Rennes, INRIA, CNRS, IRISA, France

s.djahel@mmu.ac.uk, yassine.hadjadj-aoul@irisa.fr

IEEE ISC2 2020, September 28th , 2020

2020 IEEE International Smart Cities Conference

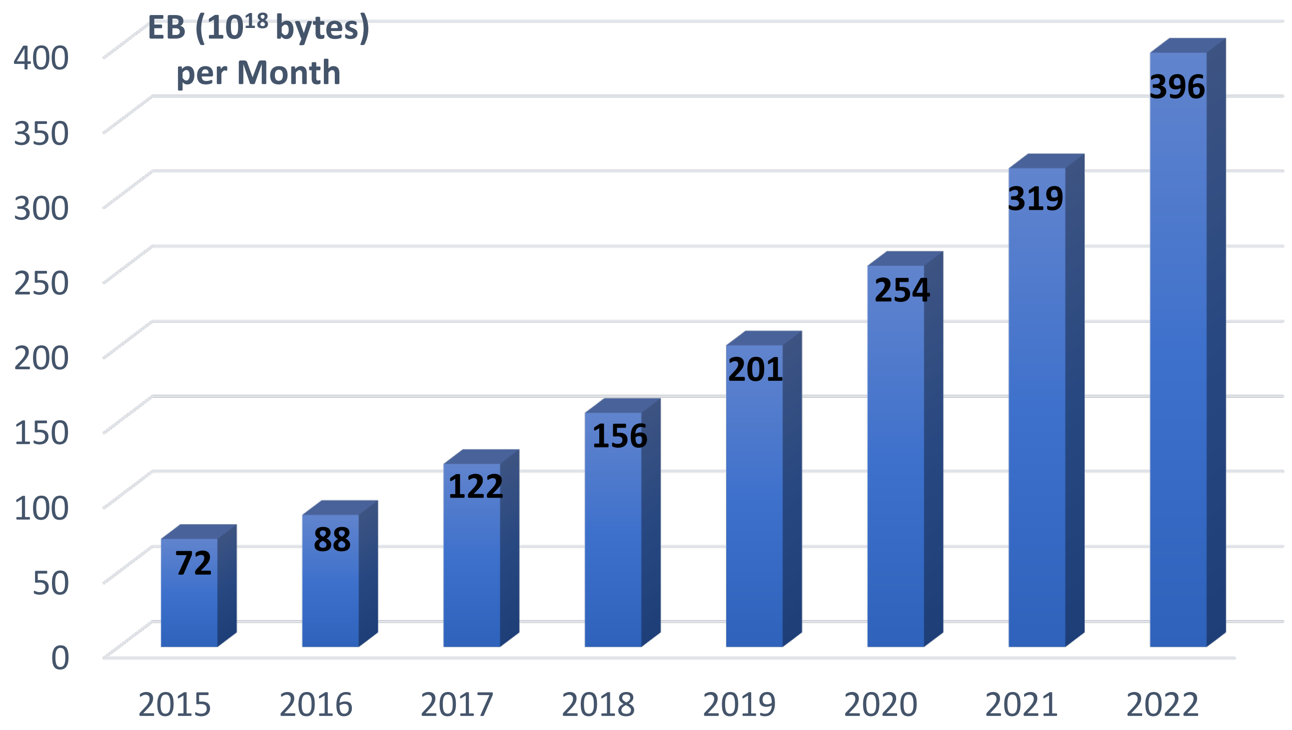
Outline

- Introduction
- Relevance of caching
- Understanding how caching works
- Caches’ placement in 5G and B5G networks
- Conclusions

Evolution of data consumption

- An endless race towards even more content and more quality
  - Data consumption increasing as never before

Cisco VNI: Global Internet traffic evolution and forecast [Cis17]



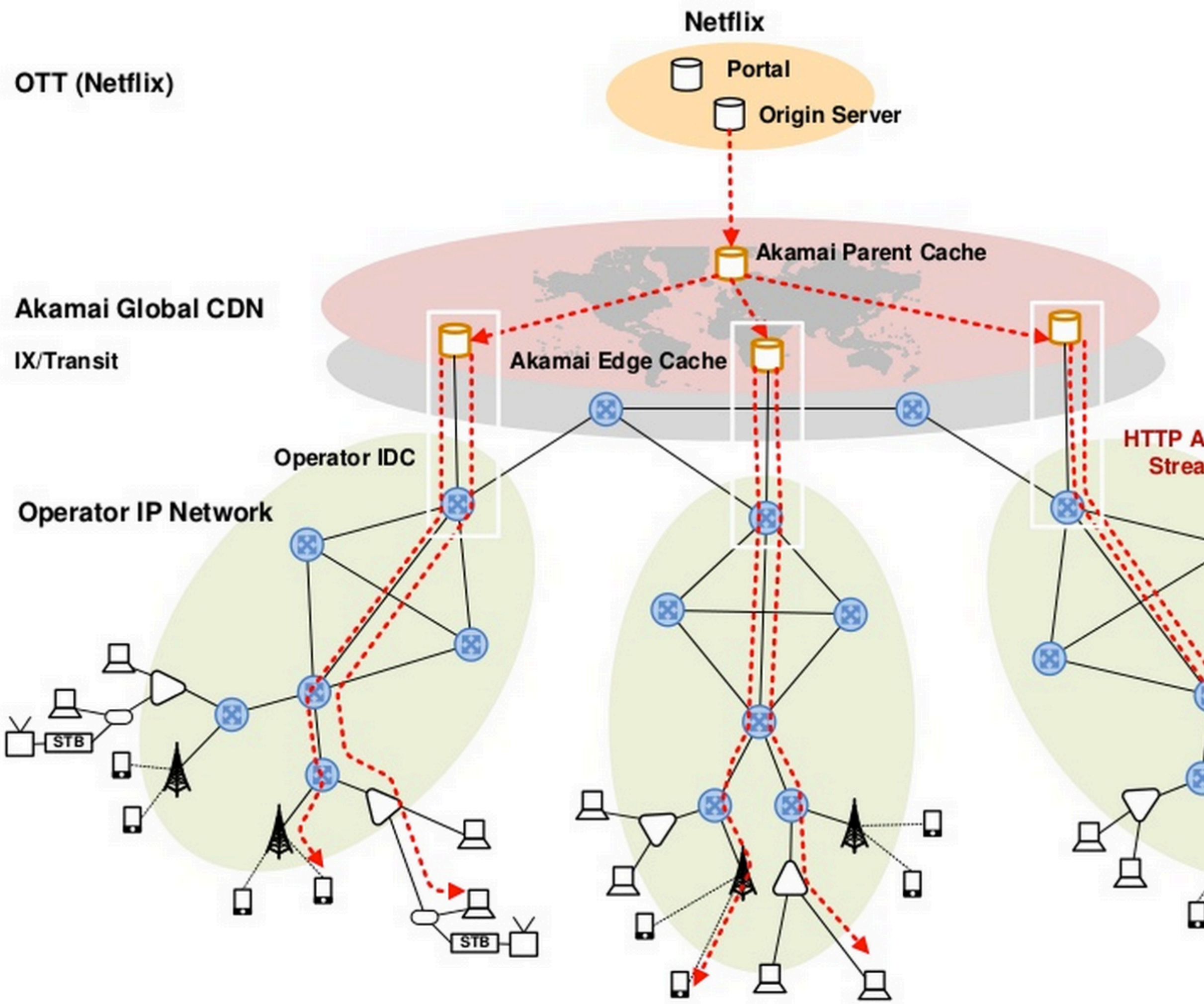
- Explosion of content retrieval and data dissemination in the Internet
- IP video will be 82% of all IP global traffic by 2022

[Cis17] Cisco, "Cisco Visual Networking Index: Forecast and methodology, 2017-2022", White paper, 2017.

Modern ecosystem (1/3)

- This growth impacts the infrastructure of ISPs and CPs/SPs
  - Congestion at the access (Edge)/backhaul levels and at peering links
  - Reduced download speed when passing through the origin servers

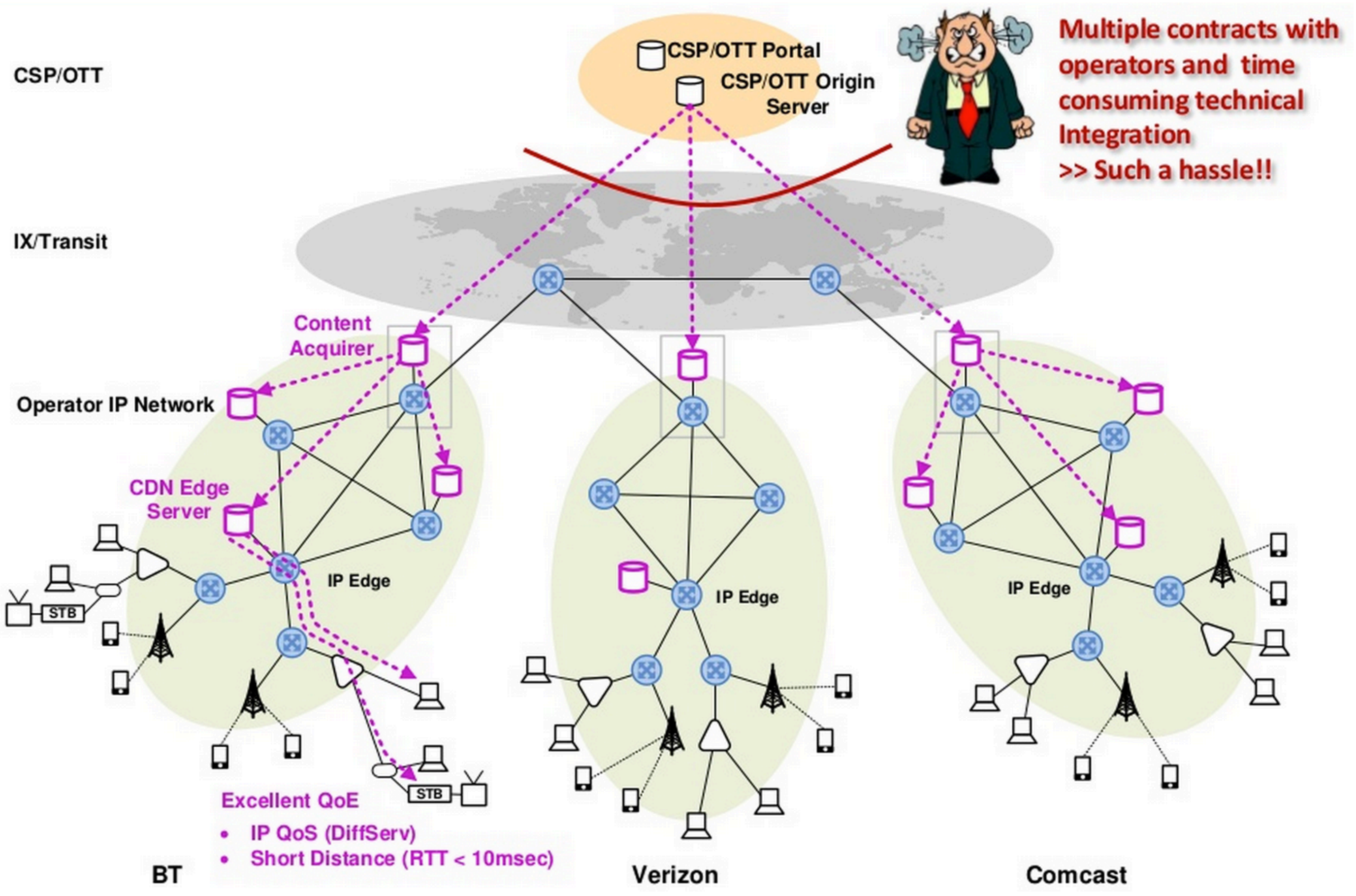
Typical CDN architecture: Hierarchical caches structure



[Adhikari12] V. K. Adhikari, Y. Guo, F. Hao, M. Varvello, V. Hilt, M. Steiner, and Z.-L. Zhang. INFOCOM, page 1620-1628. IEEE, (2012)

Modern ecosystem (2/3)

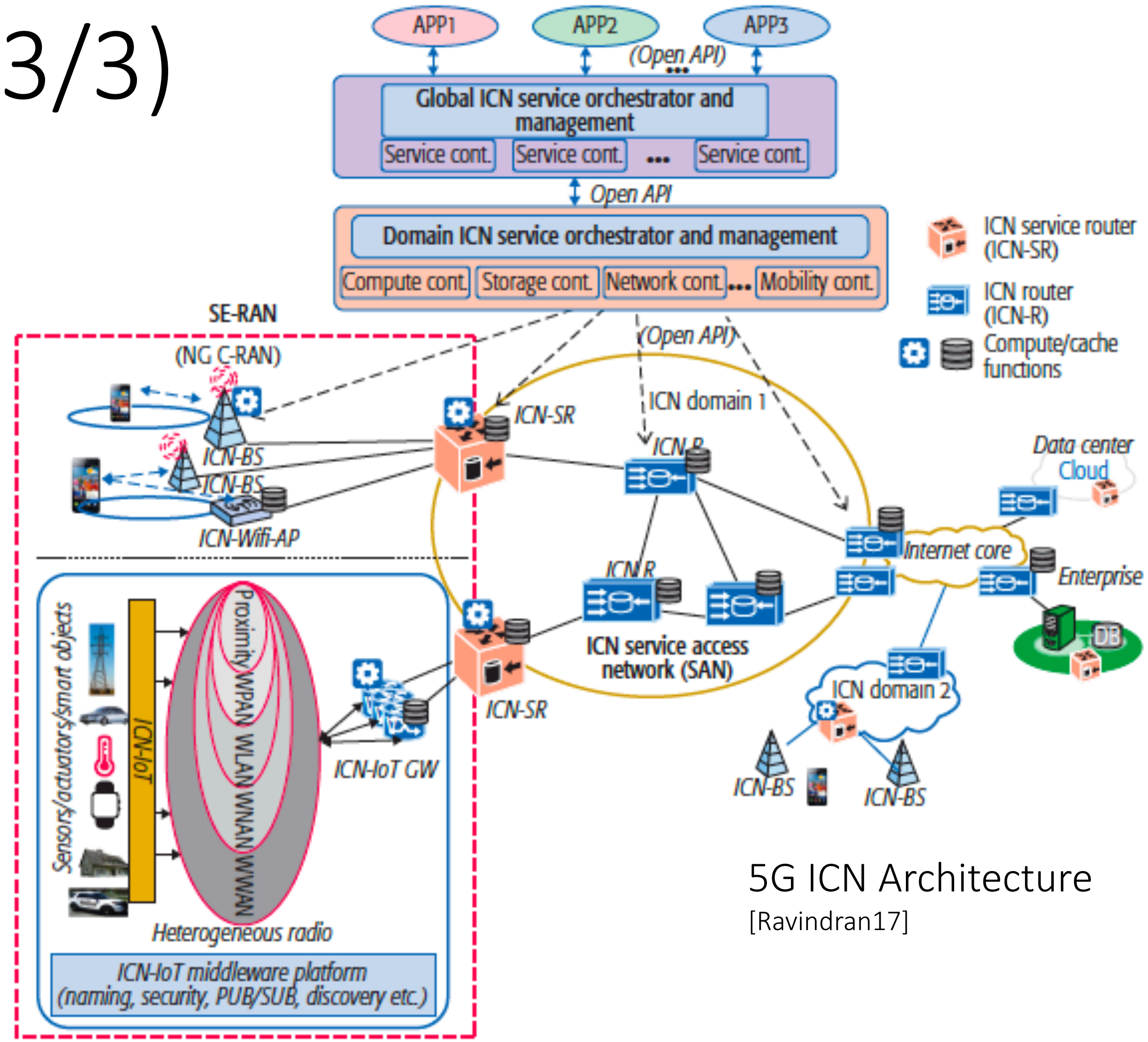
- Internet usage has evolved to become mainly used for content consumption
  - There's a need for an upgraded infrastructure better suited to today's requirements
  - Caching should go deeper into the network, to be closer to the end-users.
    - Issues with multiple operators (wholesale CDNs)



[Jac09] V. Jacobson et al., "Networking named content", CoNEXT, 2009.
[Ravindran17] R. Ravindran, et al., "5G-ICN: Delivering ICN Services over 5G Using Network Slicing," in IEEE Com. Mag., vol. 55, no. 5, pp. 101-107, May 2017

Modern ecosystem (3/3)

- Alternative approach: Content-Centric Networking (CCN) [Jac09]
  - Interesting idea but it still doesn't take ...

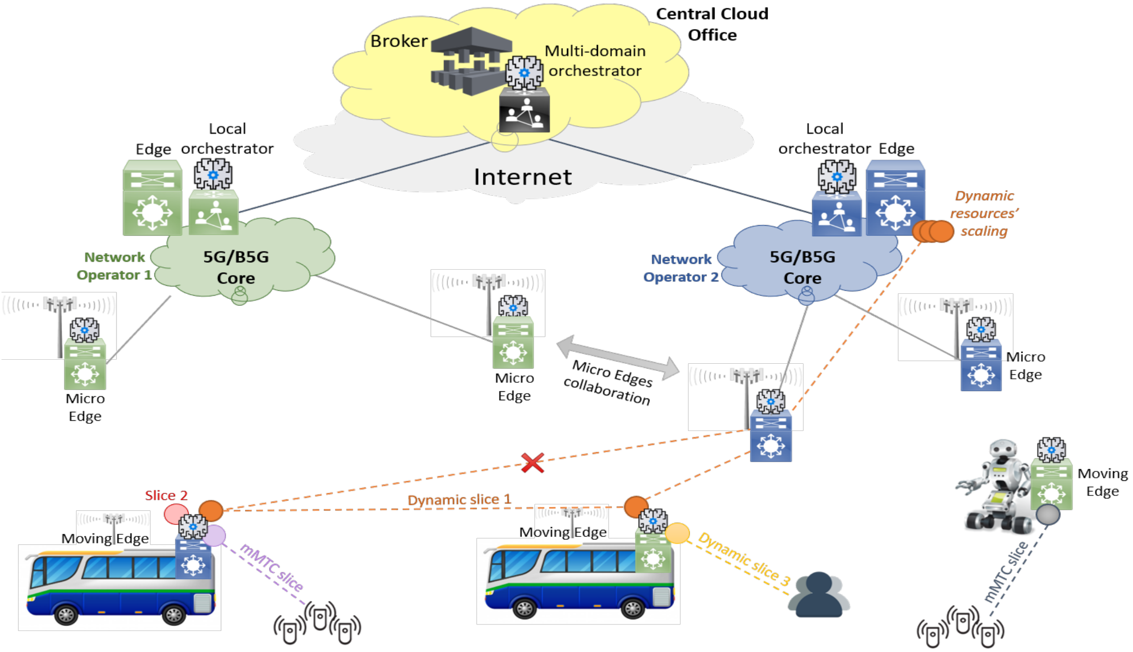


5G ICN Architecture
[Ravindran17]

[Jac09] V. Jacobson et al., "Networking named content", CoNEXT, 2009.
[Ravindran17] R. Ravindran, et al., "5G-ICN: Delivering ICN Services over 5G Using Network Slicing," in IEEE Com. Mag., vol. 55, no. 5, pp. 101-107, May 2017

New evolutions of the ecosystem (1/2)

- Evolution of the edge of the network
  - Multi-access Edge Computing (MEC) - ETSI
    - will allow cooperation between ISPs and CPs for an improved QoE
      - **this will make caching possible** (today we can only cache very few objects with the widespread use of https)
  - Potentially a mobile edge (wireless and/or ad hoc)
    - CAVs could play an important role
      - sufficient processing power is available (e.g. availability of GPUs, …)
      - potential extension of the coverage to mobile users
      - potential caching with why not a sharing of caching capabilities
        - **Everything remains to be invented …**

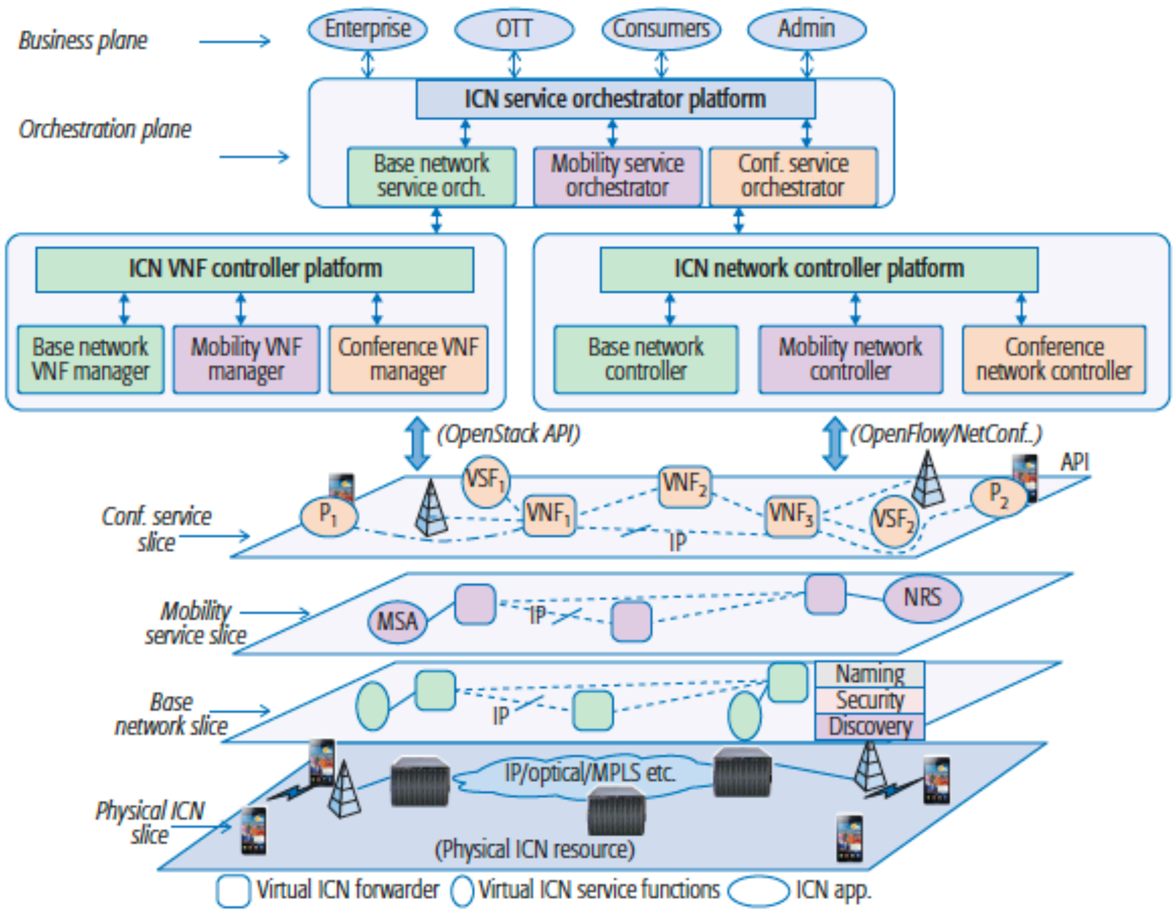


Mobile Edge Architecture

[Ravindran17] R. Ravindran, et al., "5G-ICN: Delivering ICN Services over 5G Using Network Slicing," in IEEE Com. Mag., vol. 55, no. 5, pp. 101-107, May 2017

New evolutions of the ecosystem (2/2)

- Virtualization of the network services and the infrastructures
  - Enabled by NFV and SDN
  - Possible expansion of virtualization to the field of content distribution
- 5G/B5G network slicing
  - will offer new opportunities for ISPs to dynamically manage their shared storage assets
    - Mobile edge could be included
  - ICN-as-a-slice within 5G [Rahman18]



5G ICN Architecture [Ravindran17]

[Rahman18] A. Rahman, et al., “Deployment Considerations for Information-Centric Networking (ICN)”, IETF Draft draft-irtf-icnrg-deployment-guidelines-01 (May 2018)

[Ravindran17] R. Ravindran, et al., "5G-ICN: Delivering ICN Services over 5G Using Network Slicing," in IEEE Com. Mag., vol. 55, no. 5, pp. 101-107, May 2017

The question of caches' placement becomes central

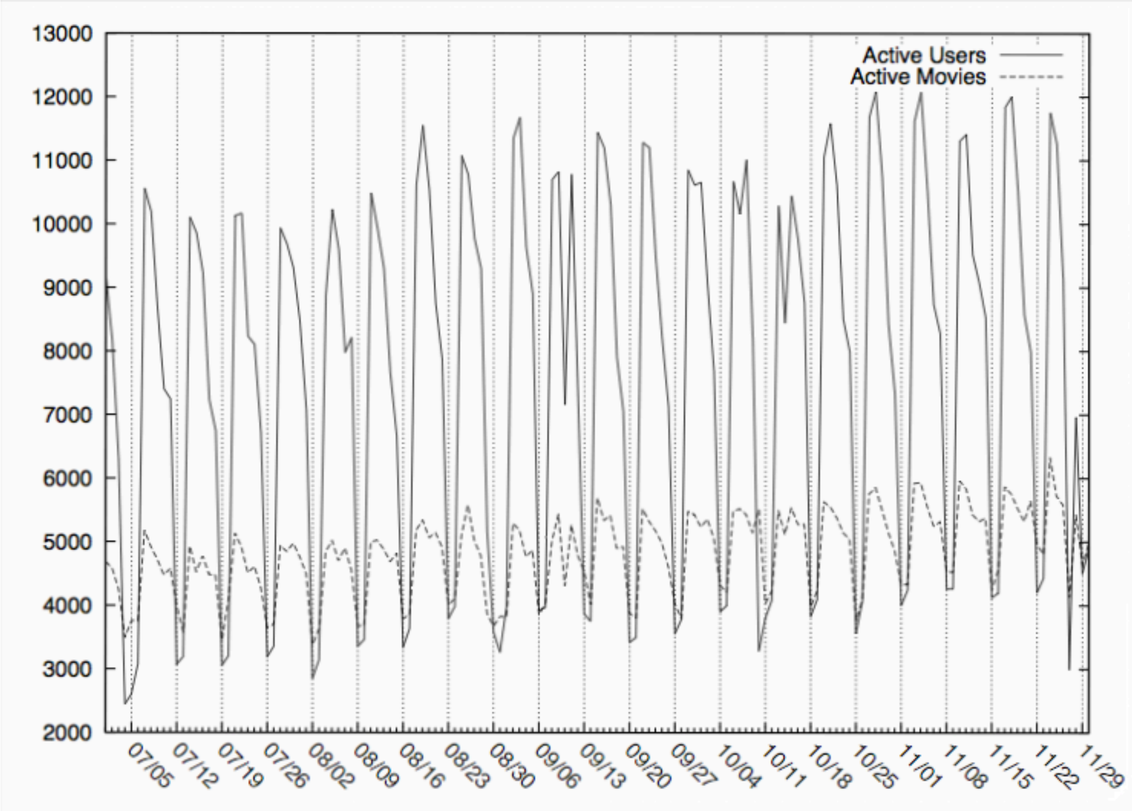
Before placing caches, let's ask proper questions ...

- Is caching really useful?
- How to measure the performance of a cache or a network of caches?
- How to place caches?

Relevance of caching

Relevance of caching for VoD

- Number of active users > Number of actives videos [Dario13]
    - Limited size of the video library
    - Popularity Zipf (α = 1.2) [Fricker12]
        - Rapid decrease of the popularity
- The caching of popular VoD content is effective



Year: 2001

[Dario13] W. Bellante, R. Vilardi and D. Rossi, "On Netflix catalog dynamics and caching performance," 2013 IEEE CAMAD, Berlin, 2013, pp. 89-93

[Fricker12] C. Fricker, P. Robert, and J. Roberts: « A versatile and accurate approximation for LRU cache performance ». In Journal Corr, Vol. abs/1202.3974 (2012)

Relevance of caching for UGC

- ANR ViPeer [Li12]
  - **Partners:** INRIA, TB, Orange, …
  - **Objective:** Relevance of caching
  - **Type of trafic:** Youtube (Almost unlimited video library)
  - **Considered ADSL zones:** Lyon, Bordeaux, and Paris
  - **Hypothesis:** the files viewed 1 and 2 times are considered as noise for caching

	Distinct files	# downloads	Total volume
Day 7			
Files	104548	223422	6493,1 GB
>2	9420	115516 (51,7%)	2757,9 GB
1		82350 (36,8%)	2634,6 GB
1 week			
Files	340702	836236	12484,6 GB
>2	43480	492482 (58,9%)	6124,7 GB
1		250690 (29,9%)	4463,6 GB
2 weeks			
Files	865681	2838514	22632,5 GB
>2	145724	1992348 (70,2%)	11916,9 GB
1		593721 (20,9%)	7483,1 GB

*[Li12] Zhe Li, Mohamed Karim Sbai, **Yassine Hadjadj-Aoul**, Annie Gravey, Damien Alliez, Jeremie Garnier, Gérard Madec, Gwendal Simon, Kamal Singh: " Network Friendly Video Distribution ". In proc. of IFIP/IEEE NoF, International Conference on Network of the Future. Tunis, Tunisia (November 2012)*

Relevance of caching for UGC

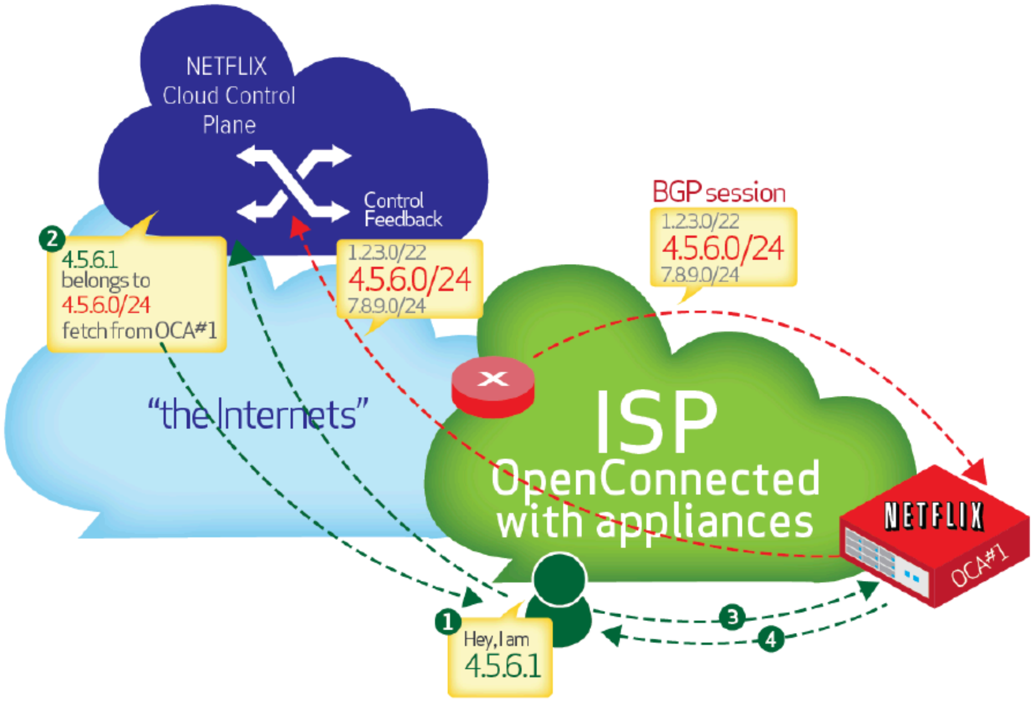
- Results (3 zones)
  - Traffic of the files viewed only 1 time is very important
  - Enlarging the study window allows to increase the number of files viewed more than 2 times (≈70%)
    - The potential of bandwith's reduction is very high
    - A cache size of 12 TB could be enough

	Distinct files	# downloads	Total volume
Day 7			
Files	104548	223422	6493,1 GB
>2	9420	115516 (51,7%)	2757,9 GB
1		82350 (36,8%)	2634,6 GB
1 week			
Files	340702	836236	12484,6 GB
>2	43480	492482 (58,9%)	6124,7 GB
1		250690 (29,9%)	4463,6 GB
2 weeks			
Files	865681	2838514	22632,5 GB
>2	145724	1992348 (70,2%)	11916,9 GB
1		593721 (20,9%)	7483,1 GB

[Guillemin13] F. Guillemin, B. Kauffmann, S. Moteau and A. Simonian, "Experimental analysis of caching efficiency for YouTube traffic in an ISP network," Proc. of the 2013 25th Int. Teletraffic Congress (ITC), Shanghai, 2013, pp. 1-9

Is caching really useful?

- Clearly beneficial for all parties:
  - Customers: we improve their quality of experience
  - CPs/SPs: their customers will be satisfied and their origin servers less solicited
  - ISPs: less congested
- However,
  - Mandatory cooperation between CPs/SPs and ISPs
    - Already done (Open connect appliance, Netflix) [Bottger18]

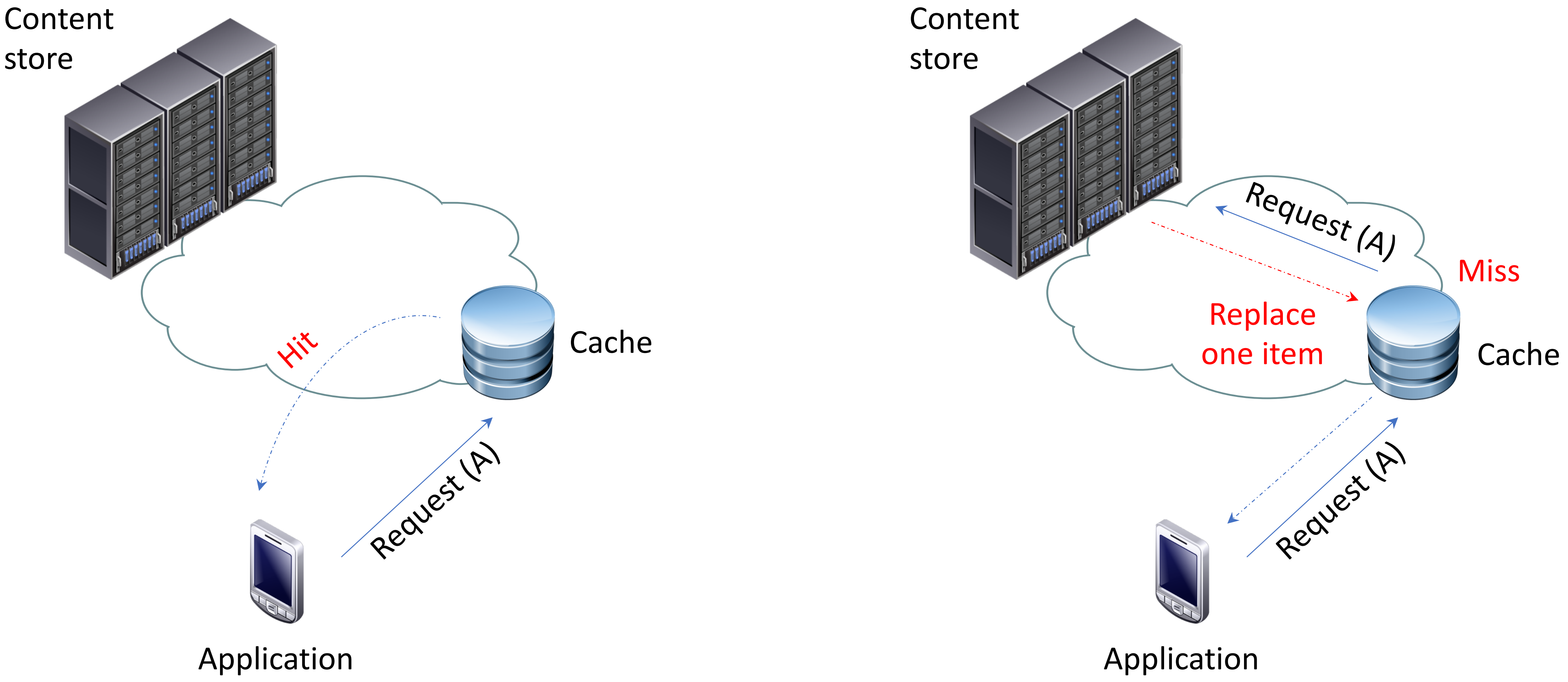


[Bottger18] T. Bottger et al. , « Open Connect Everywhere: A Glimpse at the Internet Ecosystem through the Lens of the Netflix CDN ». ACM SIGCOMM Computer Communication Review, (April 2018)

How a cache works?

Focus on a single cache

How a cache works?

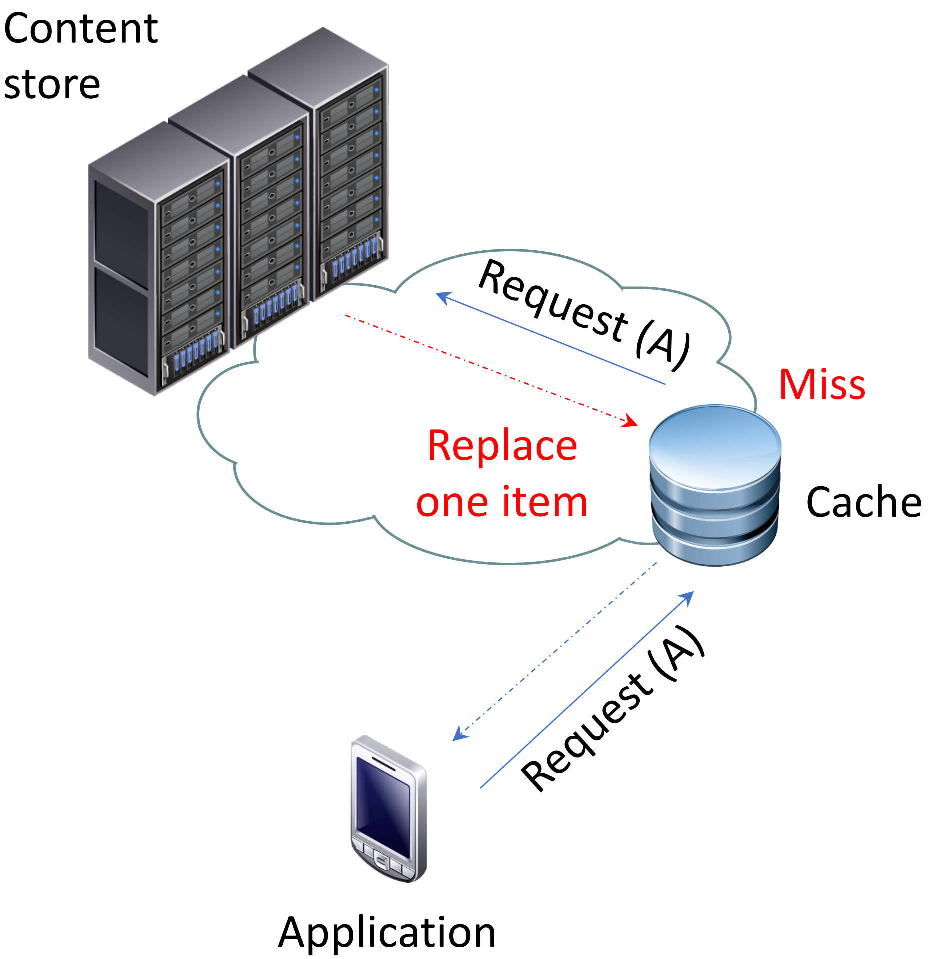


Important performance metric will be the **hit probability**

$$\text{Hit probability} = \frac{\text{number of items served from cache}}{\text{total number of items served}}$$

- Goal: find a policy to maximize the hit probability.

Which item to replace?

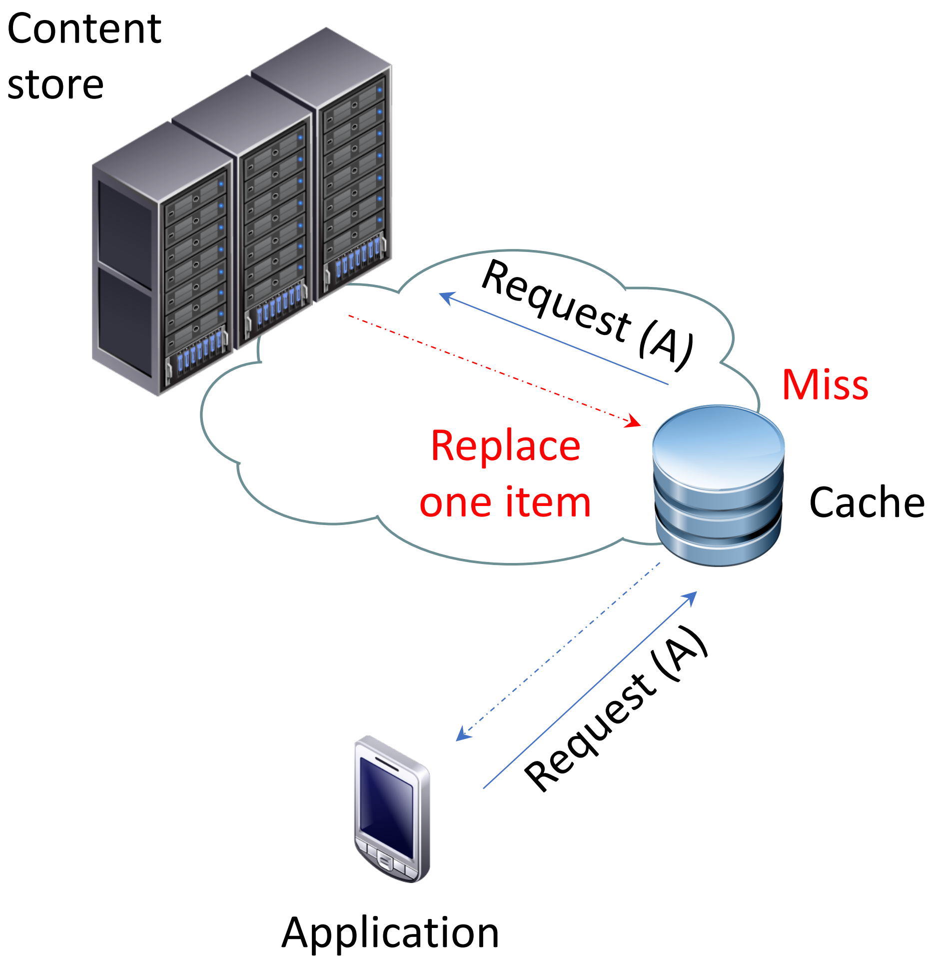


- Classical cache replacement policies:
  - RAND
  - FIFO
  - <u>LRU</u>
  - LFU
  - LRU-K

Even when the popularity is constant, LFU is not optimal.

- LFU is optimal under IRM (it maximizes the steady-state hit probability).
  - LFU statically places in the cache the C most popular objects (on the basis of the average request rate of contents)
  - Higher complexity than LRU
- LFU is no longer optimal under a general distribution
  - The content of the cache is never adapted to instantaneous traffic conditions, resulting into suboptimal performance.

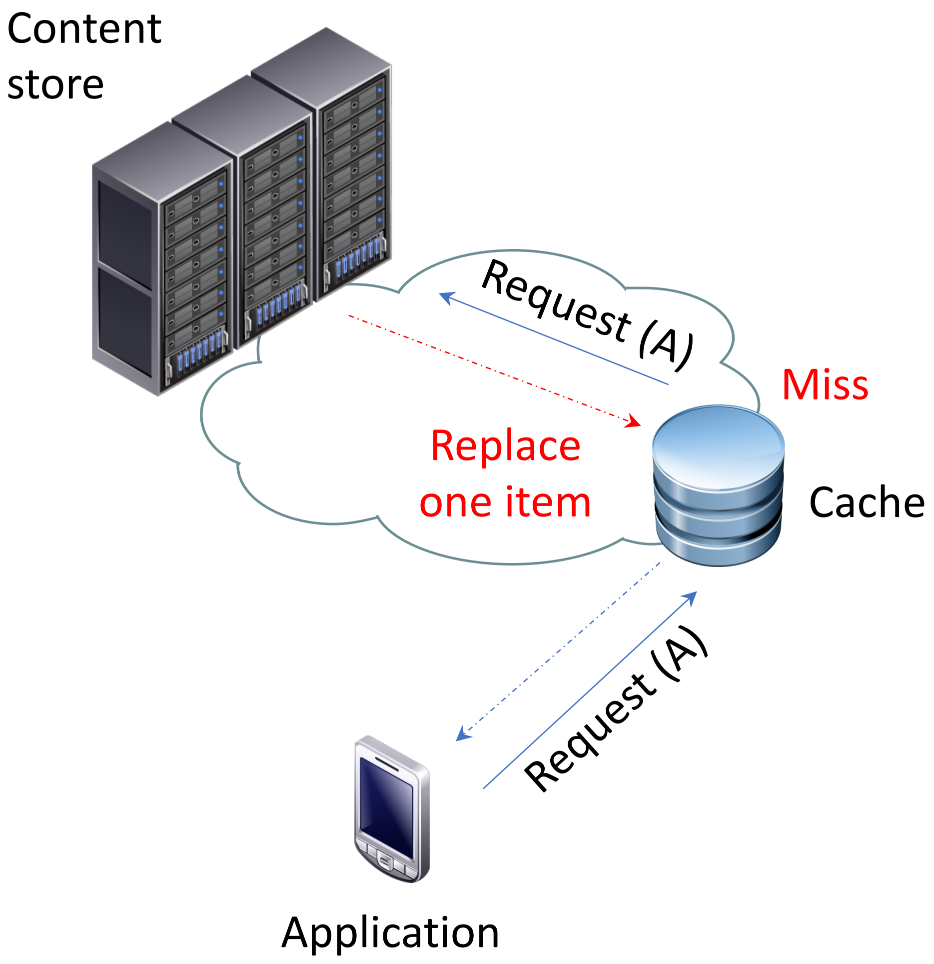
The offline problem is easy …



If you know the sequence of requests (not realistic):

- Min policy
  - At time t, if X_t is not in the cache, evict an item in the cache whose next request occurs furthest in the future.
- Theorem (Maston et al. 1970)
  - MIN is optimal

The online problem is not easy …



LRU is motivated by the heuristic that the element that has not been used for the longest time is likely to also not be needed for the longest time.

heuristic works well even in a worst-case analysis

How to compare different policies?

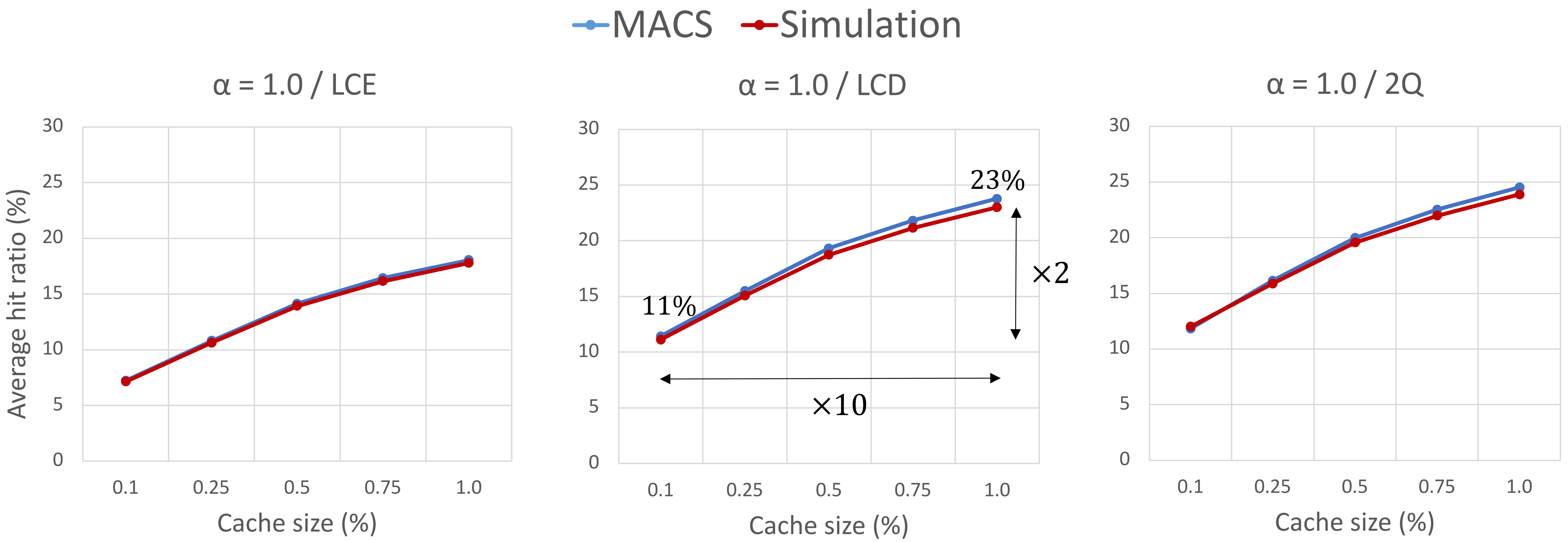
- Case 1: We can use trace-based simulations/experimentations
  - To compensate the lack of “traces”, we can model requests arrivals as stochastic processes
- Case 2: We can use models (faster) – Only the control plane is considered
  - The exact model is not tractable [King71]
  - Approximation of hierarchical caches under LRU/LCE [Che02]
  - Approximation of hierarchical caches under LRU/* (MACS) [BenAmmar19]

[King71] W.F. King III, “Analysis of demand paging algorithms”, Proc. of the IFIP Congress 1971, North-Holland Publishing Company.

[Che02] Che, H., Tung, Y., & Wang, Z. (2002). Hierarchical web caching systems: Modeling, design and experimental results. Selected Areas in Communications, IEEE Journal on, 20(7), 1305-1314.

*[BenAmmar19] H. Ben Ammar, **Y. Hadjadj-Aoul**, G. Rubino, and S. Ait Chellouche : “On the performance analysis of distributed caching systems using a customizable Markov chain model”. In Elsevier, Journal of Network and Computer Applications (JNCA). (January 2019)*

Average hit ratio of the network using MACS



- Good accuracy in estimating the overall cache hit performance (even when considering a multi-root tree structure)
- The cache becomes less efficient as its capacity increases

*[BenAmmar19] H. Ben Ammar, **Y. Hadjadj-Aoul**, G. Rubino, and S. Ait Chellouche : "On the performance analysis of distributed caching systems using a customizable Markov chain model". In Elsevier, Journal of Network and Computer Applications (JNCA). (January 2019)*

How to compensate the lack of “traces”?

- We can model requests arrivals as stochastic processes (Started with [King 1971, Gelenbe 1973])
  - Independent reference model (IRM)
    - Memoryless model
    - At each time step, item i is requested with probability p_i .
  - IRM is valid for web-caching [Shenker99]

[Shenker99] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker. Web caching and Zipf-like distributions: Evidence and implications. In INFOCOM’99, volume 1, pages 126-134. IEEE, 1999.

Which probabilities to use for the contents' request?

- Zipf distribution has been widely tested against the distributions collected from the real traces
- The probability to request the content of rank r is

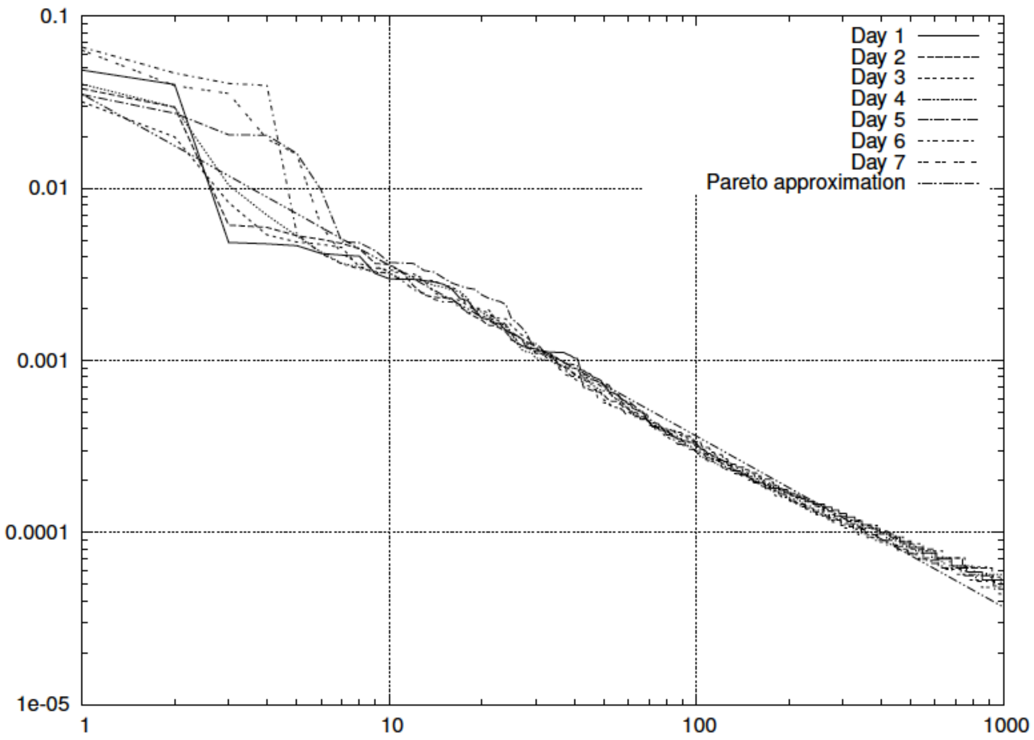
$$p_r = \frac{r^{-\alpha}}{\sum_{i=1}^{R} i^{-\alpha}}$$

- where α is the skew of the distribution; It depends on the type of the accessible objects

Zipf law

- Popularity curve of Youtube traffic []Guillemin13]

Log(frequency)



Log(rank)

Global measure for Youtube in France (ViPeer project)

More generally:

- $\alpha \sim 0.8$ for UGC traffic
- $\alpha \sim 1.2$ for VoD traffic

[Guillemin13] F. Guillemin, T. Houdoin, and S. Moteau, "Volatility of YouTube content in Orange networks and consequences," in Proc. ICC'13, Budapest, June 2013.

What is the impact of caching within a wireless and mobile edge (e.g. CAVs)?

- When considering only the control plane:
  - Mobility pattern could be considered as it impacts the traffic pattern
  - Future requests' prediction (prefetching) could improve the performance (cooperation between CAVs, and CAVs-Infrastructure) [Ghada16]
- When considering the data plane:
  - Link state variation can cause the user to choose a better or worse quality, which directly impacts the caching strategies (and the control plane)
  - The links congestion impacts the QoE [Rozhnova15]

*[Ghada16] Ghada Moualla, Pantelis A. Frangoudis, **Yassine Hadjadj Aoul**, Soraya Ait Chellouche: « A Bloom-Filter-based socially aware scheme for content replication in mobile ad hoc networks ». CCNC 2016*

[Rozhnova15] Natalya Rozhnova, « Congestion control for Content-Centric Networking ». Pierre and Marie Curie University, Paris, France, 2015

Cache placement in 5G and B5G

Network of caches

Cache placement in 5G and B5G (1/2)

- ISPs are developing more and more caching capabilities:
  - To relieve the load on their networks and improve the QoE of users

- Key questions:
  - Where to strategically place caches for a network operator?
  - For a cache budget requested by the content provider, where should the resources be reserved?

Cache placement in 5G and B5G (2/2)

- Contradictory results in the state of the art [Sah17]:
  - Caching at the edge is more effective
  - Caching at the core of the network is better

- Major drawbacks of current solutions dealing with the cache allocation problem [Sah17]:
  - Only one single metric or objective is used to place the cache

[Sah17] J. Sahoo et al., ``A Survey on Replica Server Placement Algorithms for Content Delivery Networks'', IEEE Comm. Surveys Tutorials, 2017.

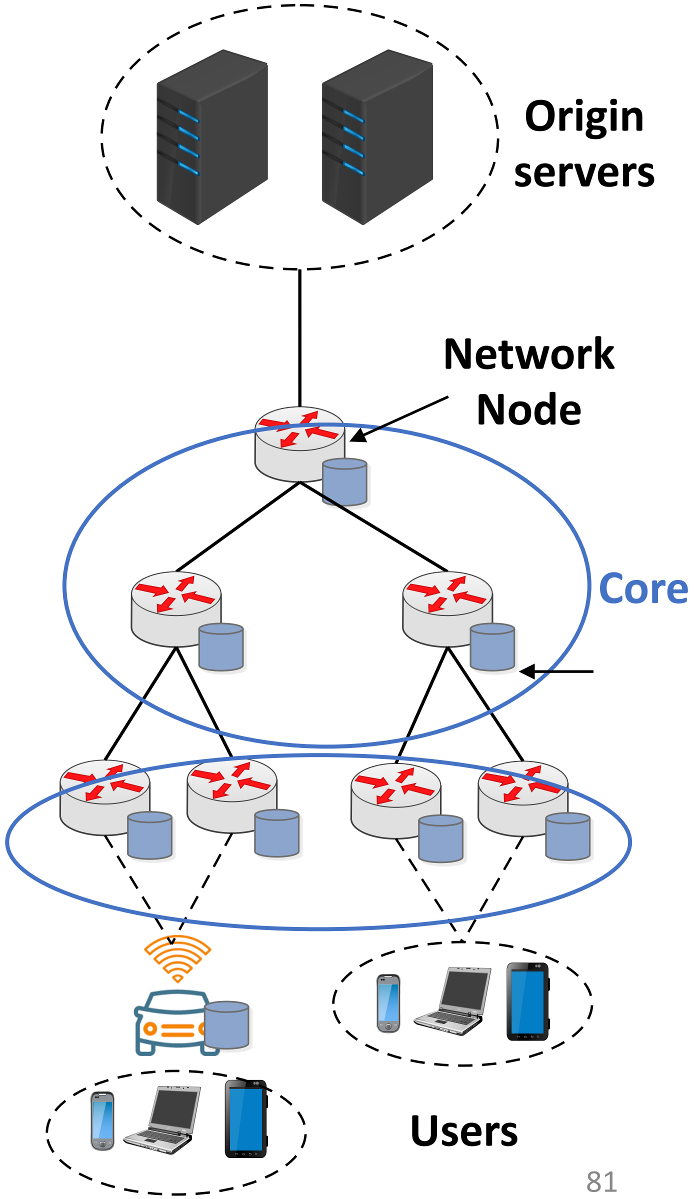
Cache placement: challenges

- Assume we have a fixed amount of cache to distribute over a multi-cache network

- Placing most of the cache at the core:
  - Allows to reduce the solicitation of origin servers (use less peering links)
  - Does not solve the problem of congestion within the ISP network

- Distributing the cache at the network edge:
  - Minimizes the delay for the most popular content
  - Less effective caching due to the content redundancy (higher demands on the origin servers)

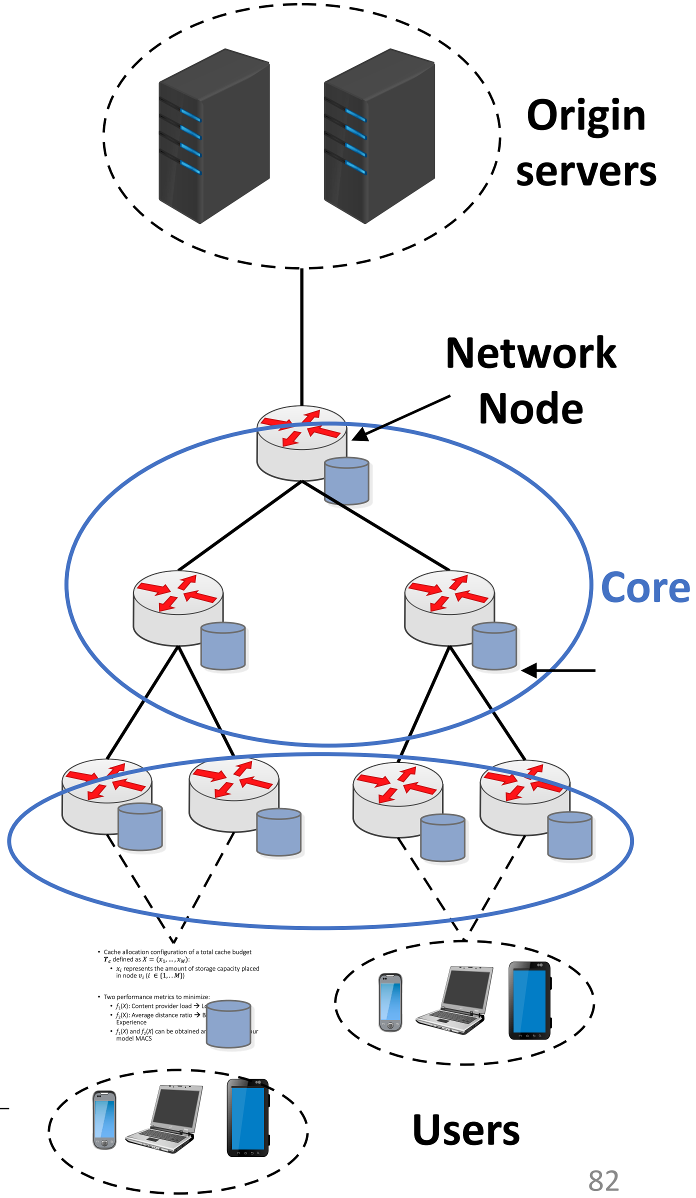
- **How to allocate efficiently the resources of cache?**



Setting up an efficient caching system (1/2)

- Cache allocation configuration of a total cache budget $\boldsymbol{T_c}$ defined as $X = (x_1, \dots, x_M)$:
  - x_i represents the amount of storage capacity placed in node v_i ($i \in \{1, \dots M\}$)

- Two performance metrics to minimize:
  - $f_1(X)$: Content provider load → Lower cost
  - $f_2(X)$: Average distance ratio → Better Quality of Experience
  - $f_1(X)$ and $f_2(X)$ can be obtained analytically using our model MACS



[BenAmmar2020] H. Ben Ammar, Y. Hadjadj Aoul: A GRASP-Based Approach for Dynamic Cache Resources Placement in Future Networks. J. Network Syst. Manage. 28(3): 457-477 (2020)

Setting up an efficient caching system (2/2)

- Multi-objective optimization problem [Ben18]:

$$\text{minimize} \qquad f_1(X), f_2(X)$$

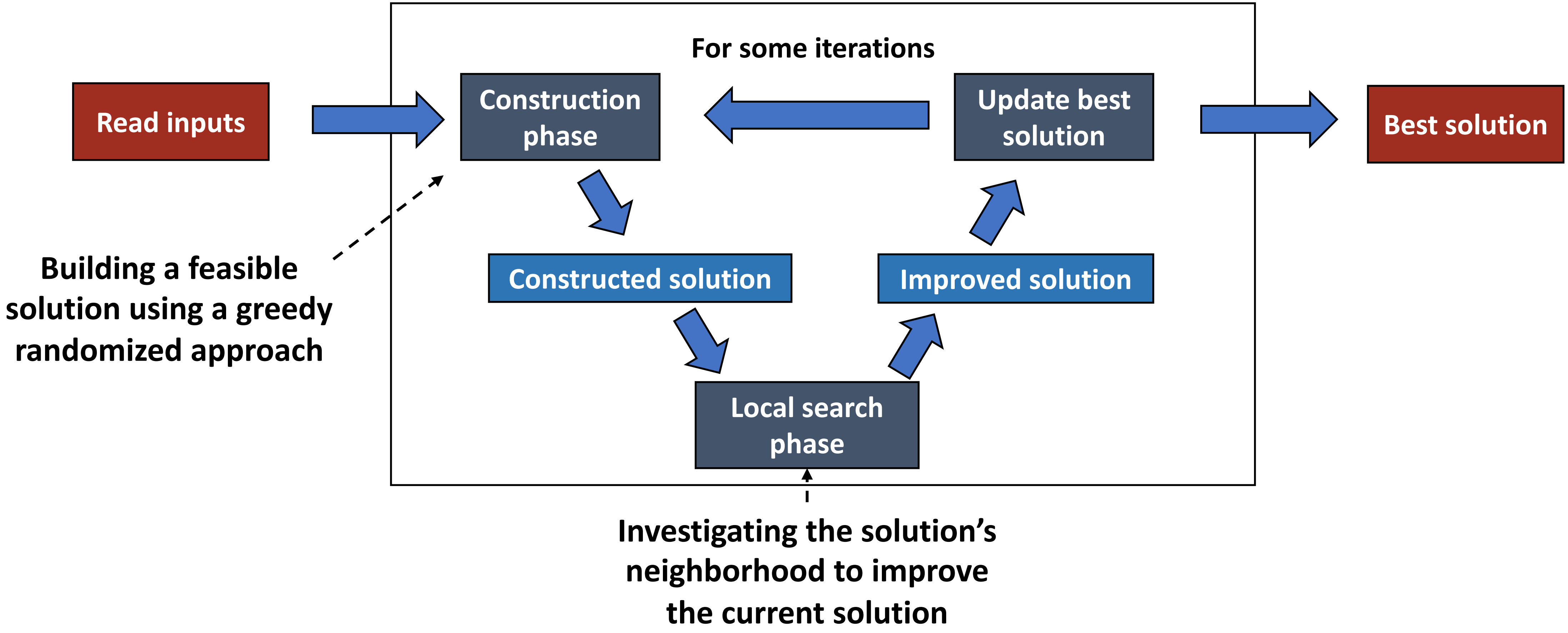
$$\text{Subject to} \qquad \sum_{i=1}^{M} x_i \leq T_c, i = 1, \ldots, M$$
$$x_i \in \mathbb{N}, i = 1, \ldots, M$$

- *Comes down to computing the weak composition* of an integer T_c into M parts
- Problem formulated as an integer nonlinear program
  - NP-hard problem

[Ben18] H. Ben Ammar, Y. Hadjadj-Aoul, "A GRASP-Based Approach for Dynamic Cache Resources Placement in Future Networks". J. Network Syst. Manage. 28(3): 457-477 (2020).

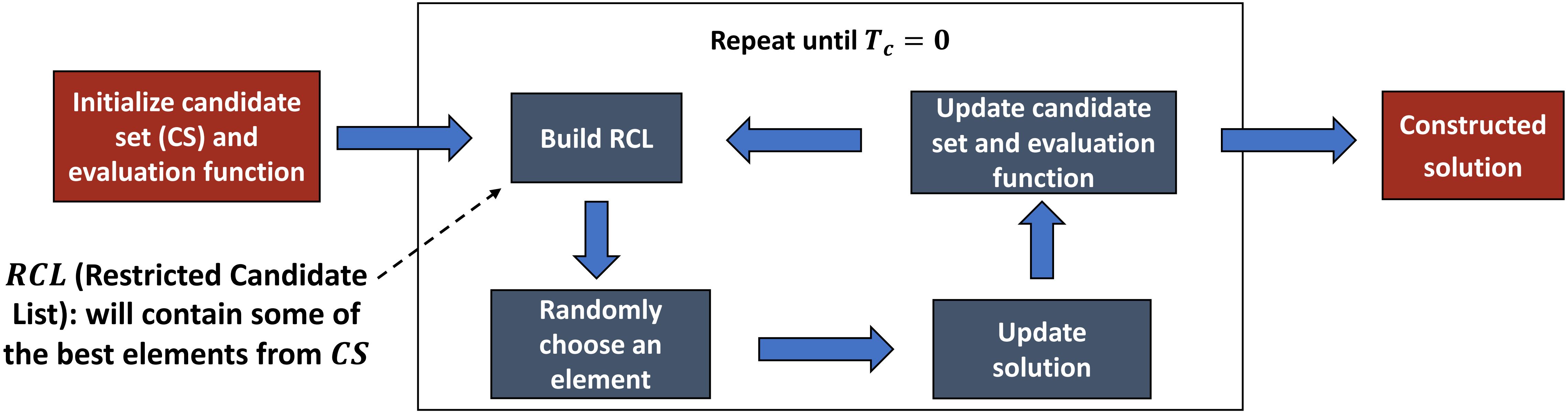
Solving the cache allocation problem using GRASP

- Using Greedy Randomized Adaptive Search Procedure (GRASP) [Feo95]
  - An iterative metaheuristic algorithm commonly applied to combinatorial optimization problems



[Feo95] T. Feo et al., ``Greedy Randomized Adaptive Search Procedures'', Journal of global optimization, 1995.

GRASP: construction phase



$X = (x_1, x_2, x_3)$
$T_c = 100$
$S = (0,0,0)$

Partial cache: $P_c = 10$
$CS = \{(10,0,0), (0,10,0), (0,0,10)\}$
$f\big((10,0,0)\big) = val_1$
$f\big((0,10,0)\big) = val_2$
$f\big((0,0,10)\big) = val_3$

$RCL \leftarrow \{X \in CS \,|\, f(X) \leq f^{min} + \lambda(f^{max} - f^{min})\}$
$\lambda \epsilon [0,1]$ ($\lambda = 0$: purely greedy / $\lambda = 1$: purely random)
$RCL = \{(10,0,0), (0,10,0)\}$
$S = (0,10,0)$
$CS = \{(10,10,0), (0,20,0), (0,10,10)\}$

GRASP: local search phase

Repeat until no better solution is found

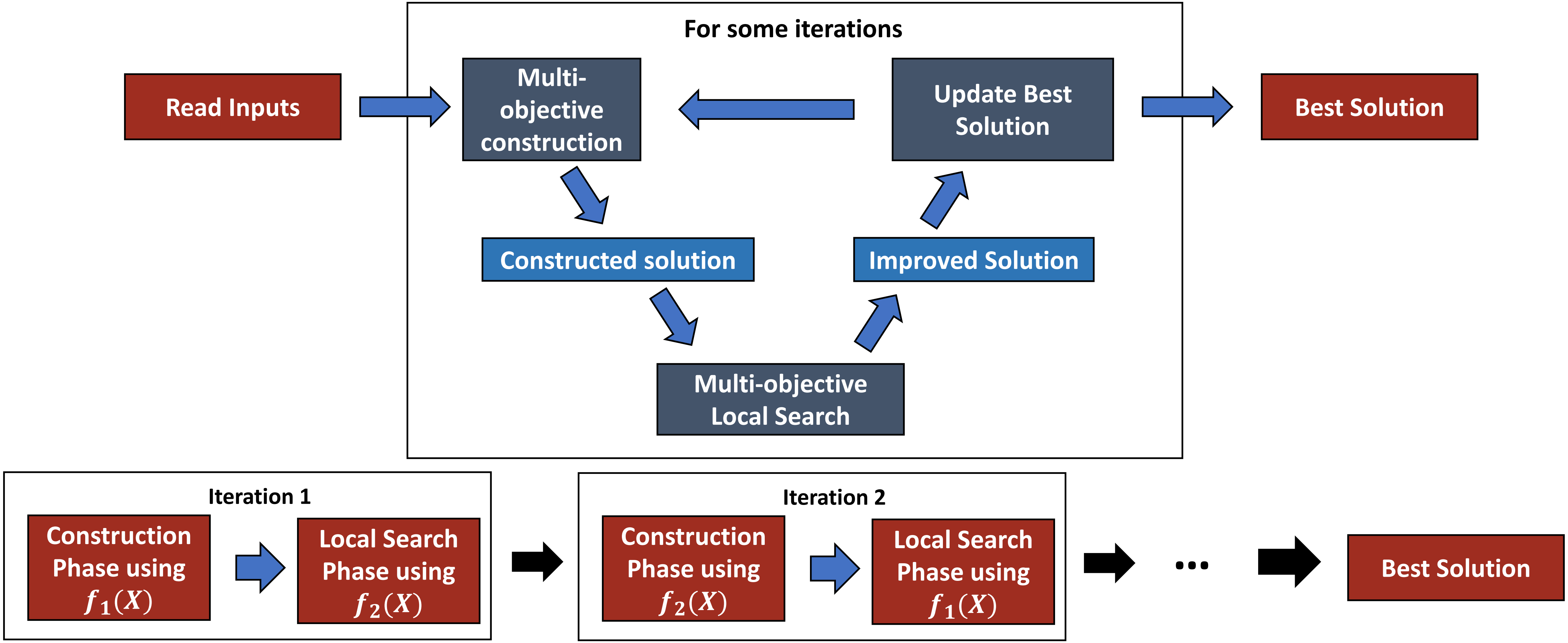
Contructed Solution → Create neighborhood → Select first-improving/best-improving solution → Update current solution → Create neighborhood

Update current solution → Improved Solution

$$S = (20,40,40)$$
$$N(20,40,40) = \{(10,50,40), (10,40,50), (30,30,40), (20,30,50), (30,40,30), (20,50,30)\}$$

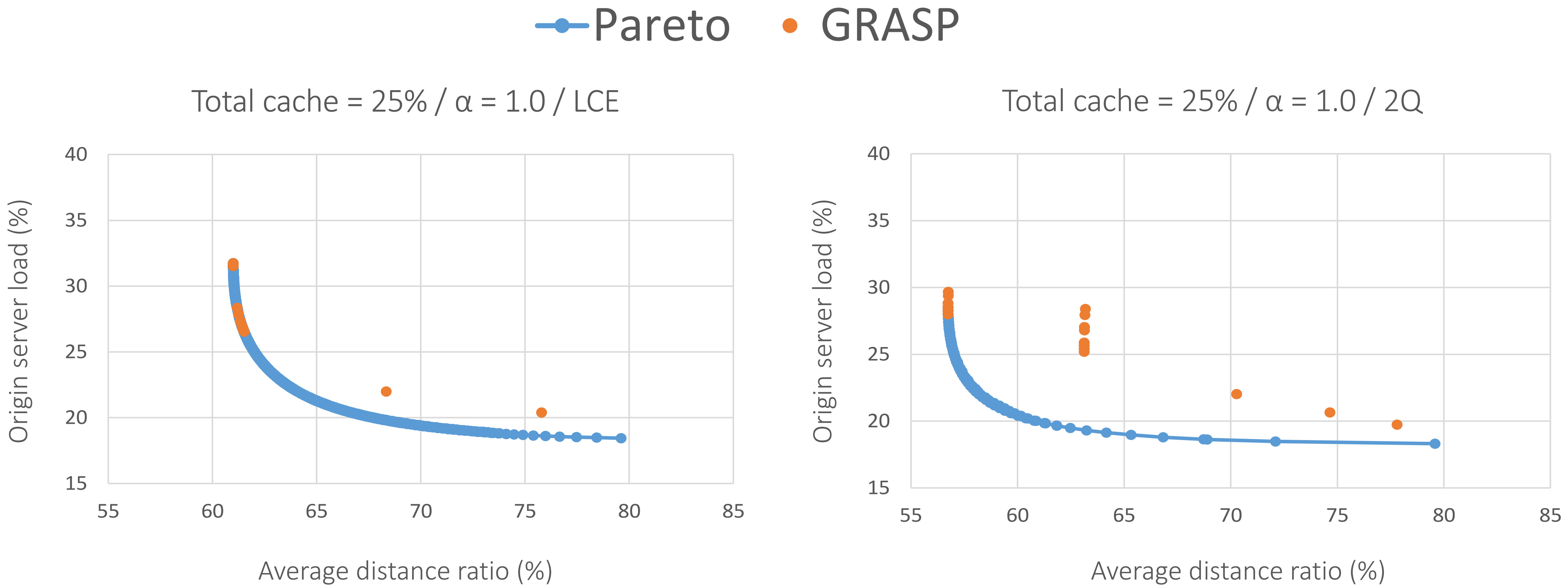
$$S = (30,30,40)$$
$$N(30,30,40) = \{(20,40,40), (20,30,50), (40,20,40), (30,20,50), (40,30,30), (30,40,30)\}$$

Multi-objective GRASP



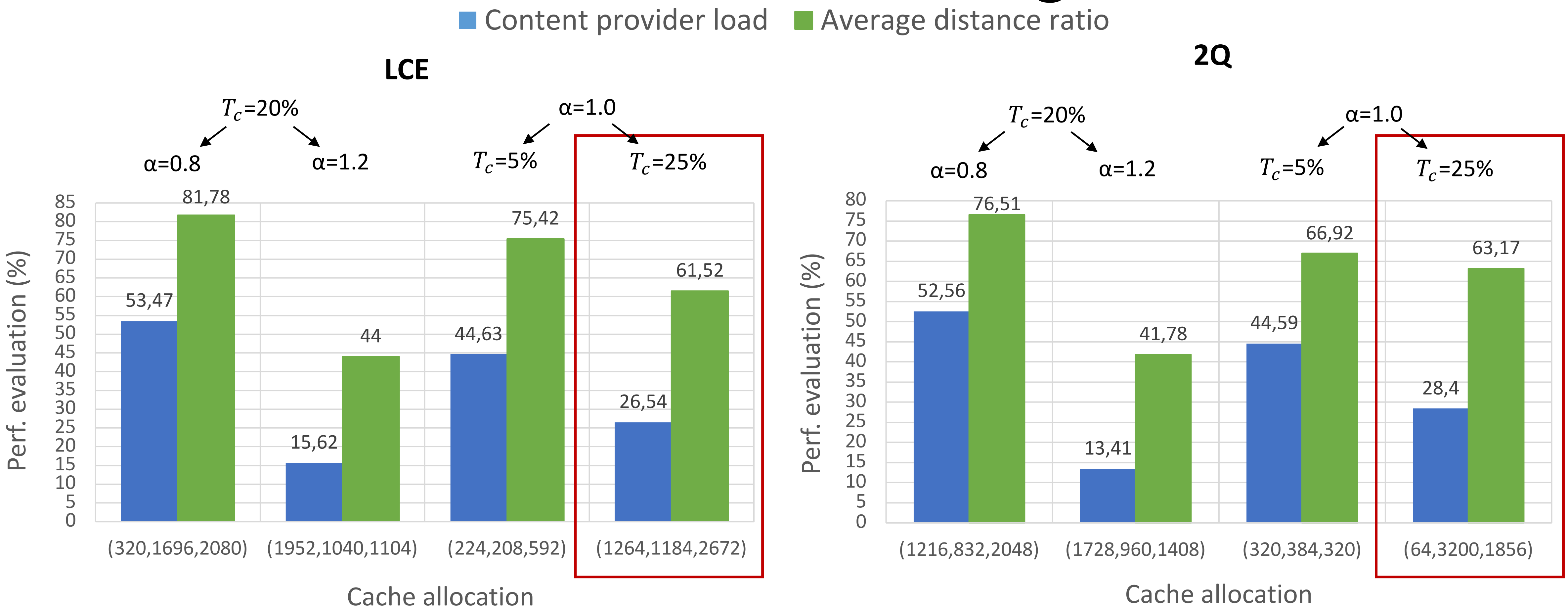
[Mar14] R. Marti et al., ``Multiobjective GRASP with Path Relinking'', European Journal of Operational Research, 2014.

GRASP solutions compared to the Pareto front



- Solutions produced by GRASP are close to the set of dominant points

Cache allocation solutions using GRASP



- The choice of a specific cache allocation depends on specific desired objectives
- A wrong placement of cache can decrease the efficiency of the caching scheme

Challenges

- Improving cache placement in 5G and B5G using real-world infrastructures
  - Deep learning could be efficient to solve such a problem
- Placing caches in the context of multi-tenant networks
  - Problem even harder (i.e., different traffic patterns, priorities, …)
- Placing in a multi-domain context
  - To solve the issue of cooperation between several operators
- Placement caches in the mobile edge
  - Limited backhouling capacity [Zhang17]
  - Unreliable channels [Towsley20]

[Zhang17] C. Zhang, P. Ren and Q. Du, "A contextual multi-armed bandit approach to caching in wireless small cell network," 2017 9th Int. Conf. on Wireless Communications and Signal Processing (WCSP), Nanjing, 2017, pp. 1-6

[Towsley20] Paulo Sena, Igor Carvalho, Antônio J. G. Abelém, György Dán, Daniel S. Menasche, Don Towsley: Caching Policies over Unreliable Channels. WiOpt 2020: 1-8

Conclusions

- Caching is effective in improving performance given the incessant increase in demand
- Understanding the traffic pattern helps improving the effectiveness of caching strategies
- Cache placement is not a simple problem where you have to put the caches at the edge or in the core of the network
    - The distribution of caches along the infrastructure is the key to getting closer to an optimum
- Including a moving edge makes the task even more complex
- There are still many challenges that need to be addressed to make cache slicing a reality